UNIVERSITE DE LA MEDITERRANEE AIX-MARSEILLE II
FACULTE DE MEDECINE DE MARSEILLE
ECOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTE

THESE DE DOCTORAT

Présentée par

Quentin LEROY

En vue de l'obtention du grade de Docteur de l'Université de la Méditerranée
Spécialité : Maladies Transmissibles et Pathologies Tropicales

Etude par transcriptomique et génomique comparative de la pathogénicité de *Coxiella burnetii* : Une approche puce à ADN

Soutenue le Mardi 14 décembre 2010

COMPOSITION DU JURY

Pr. Jean-Louis Mège	Président du Jury
Pr. Max Maurin	Rapporteur
Pr. Marie-Laure Joly-Guillou	Rapporteur
Dr. Patricia Renesto	Examinateur
Pr. D. Raoult	Directeur de Thèse / du Laboratoire

Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes
CNRS-IRD UMR 6236

Avant propos

Le format de présentation de cette thèse correspond à une recommandation de la spécialité Maladies Infectieuse et Microbiologie, à l'intérieur du Master des Sciences de la Vie à Marseille. Le candidat est amené à respecter des règles qui lui sont imposées et qui comportent un format de thèse utilisé dans le Nord de l'Europe permettant un meilleur rangement des thèses traditionnelles. Par ailleurs, la partie introduction et la bibliographie sont remplacées par une revue envoyée dans un journal afin de permettre une évaluation extérieure de la qualité de la revue et permettre à l'étudiant de commencer le plus tôt possible une bibliographie exhaustive sur le domaine de cette thèse. Par ailleurs la thèse est présentée sur article publié, accepté ou soumis associé d'un bref commentaire donnant le sens général du travail. Cette forme de présentation a paru plus en adéquation avec les exigences de la compétition internationale et permet de se concentrer sur des travaux qui bénéficieront d'une diffusion internationale.

Professeur Didier RAOULT

SOMMAIRE

Résumé

L'objectif de cette thèse a été d'enrichir nos connaissances sur les bactéries intracellulaires strictes et spécialement *Coxiella burnetii*, agent responsable de la fièvre Q. Pour ce faire, nous avons d'une part amélioré les techniques de préparation de l'ARN pour les études transcriptionnelles et d'autre part utilisé la technologie des puces à ADN pour analyser le transcriptome ainsi que la diversité génomique de *C. burnetii.*

L'utilisation d'échantillons cliniques dans les études transcriptionnelles est limitée par la quantité de matériel disponible qui ne permet pas d'analyser simultanément les profils du pathogène et de son hôte au cours de l'infection. De ce fait, nous avons développé, sur un modèle d'escarres obtenus à partir de malades de fièvre boutonneuse méditerranéenne, une stratégie basée sur l'hybridation soustractive pour séparer les ARN eucaryotes et procaryotes dans le but d'entreprendre des hybridations de puces à ADN.

C. burnetii est une bactérie hautement résistante aux stress environnementaux comme le changement de pH, la dessiccation, mais aussi le changement de température. Nous avons particulièrement étudié la réponse précoce de *C. burnetii* lors d'une exposition à une haute et faible température. L'analyse globale du profil transcriptionnel a montré que la réponse de *C. burnetii* était limitée et similaire pour les différents stress appliqués. Cependant, malgré cette faible réponse, il apparait clairement qu'une accumulation de ppGpp, un arrêt de la croissance et des modifications de la membrane et de la paroi cellulaire permettraient à *C. burnetii* de résister à ces stress. Toutes ces régulations géniques convergent vers un changement d'état de la bactérie vers une forme pseudo-sporulée. De plus, nous avons observé une organisation spatiale des gènes différentiellement exprimés. Nos analyses bio-informatiques ont montré que ces clusters de régulation ne répondaient ni au paradigme promoteur - facteur de transcription – opéron, ni à des réseaux biologiques. Ayant retrouvé ce phénomène dans plusieurs autres études transcriptionnelles chez d'autres bactéries intracellulaires, nous spéculons que ces clusters de régulations pourraient être dus à une régulation épigénétique qu'il reste à caractériser.

Différentes méthodes de typage ont déjà été mises au point pour classer les isolats de *C. burnetii* dans le but d'explorer son pouvoir pathogène. Ici, nous présentons une méthode de génomotypage basée sur la présence ou l'absence de gènes à l'aide de puces à ADN. Nous avons testé notre stratégie de génomotypage sur 52 isolats provenant de différentes zones géographiques, de différents hôtes et de patients présentant différentes manifestations cliniques. L'analyse a révélé la présence de 10 génomotypes organisés en 3 groupes avec une topologie congruente à celle observée avec le Multi Spacer Typing. Nous avons aussi découvert 4 génomotypes particulièrement associés à la fièvre Q aiguë, alors que tous les génomotypes étaient associés à la forme chronique. De plus, le génomotypage a révélé que les isolats retrouvés dans les tiques dures, y compris la souche de référence Nine Mile appartiennent au même genomotype.

Globalement, les données que nous avons obtenues confirment le fait que les puces à ADN sont un outil adapté pour l'analyse de la pathogénicité de *C .burnetti* mais aussi des autres bactéries intracellulaires strictes. Cependant, de nouvelles technologies plus résolutives comme le DNA ou RNAseq semblent être plus prometteuses mais restent encore à optimiser.

Abstract

The objective of this thesis was to increase knowledge of obligate intracellular bacteria and specifically, the causative agent of Q fever *C. burnetii.* In this regard, we have improved strategies to purify RNA for the transcriptional studies. We also used the technology microarrays to analyze the transcriptome and genomic diversity of *C. burnetii.*

The use of clinical samples in the transcriptional studies is limited by the amount of material available and thus the transcriptional profiles of the pathogen and its host during infection can not be simultaneously analyze. We developed, with a model of eschars obtained from Mediterranean spotted fever patients, a strategy based on subtractive hybridization to separate RNA eukaryotic and prokaryotic cells in order to perform microarray experiments.

Analysis of the survival strategies used by this bacterium to adapt to new environmental conditions is critical for our understanding of *C. burnetii* pathogenicity. Here, we report the early transcriptional response of *C. burnetii* under temperature stresses. Our data show that *C. burnetii* exhibited minor changes in gene regulation under short exposure to heat or cold shock. While small differences were observed, *C. burnetii* seemed to respond similarly to cold and heat shock. The expression profiles obtained using microarrays produced in-house were confirmed by quantitative RT-PCR. Under temperature stresses, 190 genes were differentially expressed in at least one condition, with a fold change of up to 4. Globally, the differentially expressed genes in *C. burnetii* were associated with bacterial division, (p)ppGpp synthesis, wall and membrane biogenesis and, especially, lipopolysaccharide and peptidoglycan synthesis. These findings could be associated with growth arrest and witnessed transformation of the bacteria to a spore-like form. Unexpectedly, clusters of neighboring genes were differentially expressed. These clusters do not belong to operons or genetic networks; they have no evident associated functions and are not under the control of the same promoters. We also found undescribed but comparable clusters of regulation in previously reported transcriptomic analyses of intracellular bacteria, including *Rickettsia* sp. and *Listeria monocytogenes*. The transcriptomic patterns of *C. burnetii* observed under temperature stresses permits the recognition of unpredicted clusters of regulation for which the trigger mechanism remains unidentified but which may be the result of a new mechanism of epigenetic regulation.

Different typing methods have been previously developed to classify *C. burnetii* isolates in order to explore its pathogenicity. Here, we report a comprehensive genomotyping method based on presence or absence of genes using microarray. The genomotyping method was then tested on 52 isolates obtained from different geographic areas, different hosts and isolated from patient with different clinical manifestations. The analysis reveals the presence of 10 genomotypes organized in 3 groups with a topology congruent with that of Multi Spacer Typing. We also found out 4 genomotypes especially associated with acute Q fever whereas all the genomotypes could be associated to chronic human infection. Serendipity, genomotyping reveals that hard ticks isolates including Nine Mile belong to the same genomotype.

Overall, the data we obtained confirm that DNA microarrays are a suitable tool for exploring pathogenicity of *C. Burnetti* and other obligate intracellular bacteria. However new technologies such as DNAseq or RNAseq seem more promising but still need to optimize and also are still expensive compared to microarray.

Introduction

Les bactéries peuvent être classées selon leur localisation en bactéries extracellulaires, intracellulaires facultatives et intracellulaires obligatoires. Les bactéries intracellulaires obligatoires n'ont pas la capacité *in vivo* de se multiplier en dehors de cellules hôtes eucaryotes. Certaines des bactéries permettent une association avantageuse pour leur hôte, définissant le mutualisme [1]. Cependant, d'autres bactéries ont un effet délétère sur leur hôte, dans ce cas on parle de parasitisme [1]. Plusieurs de ces bactéries parasites sont pathogènes pour l'Homme comme peuvent l'être les Rickettsies [2], les Chlamydia [3], les Ehrlichia [4], les Anaplasma [5], *Mycobacterium leprae [6]* en autres, mais aussi *Coxiella burnetii [7].* Ces bactéries peuvent se développer dans des compartiments cellulaires différents comme le cytoplasme et le noyau pour les Rickettsies [2] mais aussi les phagosomes ou phagolysosomes pour les Chlamydia et *C. burnetii [8]*. Ces bactéries sont à culture fastidieuse et comportent des génomes en cours d'évolution réductive et relativement petits comparés à d'autres bactéries [9–12].

Connaitre les mécanismes de régulation génétique et le polymorphisme chez ces bactéries peut permettre d'augmenter les connaissances sur la pathogénicité des bactéries en général. Pour cela, la génétique bactérienne et la création de mutants ont été largement utilisées pour comprendre comment les bactéries pouvaient interagir avec leurs cellules hôtes. Ces stratégies étaient utilisées pour identifier et caractériser des gènes associés à la virulence bactérienne [13]. Cependant, la création de mutants est difficile à réaliser pour les bactéries intracellulaires strictes, du fait de problèmes de sélection des mutants [14]. Cette problématique est à l'origine du manque de connaissance relative sur les facteurs de virulence des bactéries intracellulaires strictes. Récemment, quelques études ont été réalisées à partir de mutants pour *Chlamydia trachomatis* [15], *Rickettsia prowazekii* [16] et *C. burnetii* [17]. Mais les connaissances sur les interactions hôte-pathogène et les mécanismes de virulence de ces bactéries restent toutefois extrêmement limitées.

Lors des dix dernières années, le séquençage massif des génomes bactériens ainsi que le développement et l'amélioration de nouveaux outils robotiques ont fourni de nouvelles méthodes pour caractériser la pathogénicité des bactéries [18]. Ces méthodes dites de « Post Génomique » permettent des analyses à haut débit et génèrent d'importants volumes de données détaillées qui peuvent être utilisées pour augmenter les connaissances sur la pathogénicité des bactéries. Contrairement à la biologie moléculaire qui se limitait à l'analyse de quelques gènes [19,20], les technologies de « Post Génomique » sont capables d'englober l'ensemble d'un génome. L'une de ces technologies est la puce à ADN ou microarray [21]. Les puces à ADN reposent sur le principe d'hybridation complémentaire des acides nucléiques. Elles permettent principalement d'analyser le profil transcriptionnel d'une bactérie dans un état donné lorsque le matériel de départ est de l'ARN [22] ou bien d'analyser son contenu génique lorsque de l'ADN génomique (ADNg) est utilisé [23]. Les études de transcriptomique et de génomique comparative permettent d'identifier des gènes potentiellement responsables de la virulence bactérienne.

Dans cette thèse, en guise d'introduction, nous avons fait un état des lieux de l'ensemble des publications scientifiques concernant l'utilisation des puces à ADN pour l'étude de la pathogénicité des bactéries intracellulaires strictes. Nous avons développé une partie sur le principe général des puces à ADN et expliqué les différentes étapes suivies lors de son utilisation. De part leurs caractéristiques, les bactéries intracellulaires strictes imposent certaines limitations lors de l'utilisation des puces à ADN. Nous avons décrit ces limitations et les différentes stratégies utilisées lors des études pour les contourner. Enfin, les principaux résultats obtenus sur l'ensemble des publications, impliquant une analyse transcriptionnelle ou de génomique comparative à l'aide de puces à ADN, ont été intégrés à cette revue.

L'utilisation des puces à ADN dans le champ des interactions hôte-pathogène reste limitée [24]. A ce jour aucune étude n'a permis d'analyser simultanément les profils transcriptionnels du pathogène et de son hôte par puce à ADN à partir d'un

même échantillon clinique. En effet, la quantité limitée de ces échantillons souvent ne permet pas d'entreprendre ce genre d'expérience. Dans cette optique d'analyse complète de la réponse du pathogène et de l'hôte, nous avons développé une stratégie permettant de séparer les ARN eucaryotes et procaryotes grâce à une technique d'hybridation soustractive. Cette stratégie a été développée sur un modèle utilisant des escarres obtenues à partir de patients souffrant de fièvre boutonneuse méditerranéenne [25].

Ensuite, dans notre travail de thèse, nous nous sommes intéressés plus particulièrement à *C. burnetii*. *C. burnetii* est une petite gamma-protéobactérie Gram-négatif, identifiable à la coloration de Gimenez [7]. *C.burnetii* entre dans la cellule (monocyte ou macrophage) par phagocytose et se multiplie dans un compartiment phagosomal ou phagolysosomal à un pH de 4,5 [26]. A l'état naturel, lorsque la bactérie est isolée à partir de prélèvements animaux ou humains, elle exprime un antigène de phase I associé à un lipopolysaccharide (LPS) complet et à un potentiel d'infectiosité majeur. En laboratoire, après plusieurs passages en culture cellulaire ou sur œufs embryonnés, la perte partielle du LPS, induisant une variation antigénique, conduit à la phase II qui est non virulente [27,28]. Cette variation est associée à une délétion génomique de gènes impliqués dans la synthèse du LPS [29]. Par ailleurs, *C. burnetii* présente un cycle de développement comportant une forme intracellulaire dite « Large Cell Variant » (LCV) et une forme extracellulaire plus petite et métaboliquement inactive dite « Small Cell Variant » (SCV) [30]. Une troisième forme souvent associée à la forme SCV, caractérisée par une forme pseudo-sporulée, serait à l'origine de la résistance extrême de *C. burnetii* aux milieux extérieurs hostiles (désinfectants, pH, température, dessiccation et pression osmotique) [7,31].

C. burnetii est responsable de la fièvre Q, zoonose répandue dans le monde entier. Si l'infection par *C. burnetii* se révèle asymptomatique dans près de 60% des cas, la fièvre Q peut se manifester selon 2 formes, l'infection aiguë ou l'infection chronique [32]. La fièvre Q aiguë est principalement caractérisée par un syndrome

pseudo grippal au cours duquel peuvent notamment survenir une pneumonie ou une hépatite, alors que pour la fièvre Q chronique les patients peuvent présenter une endocardite, une infection vasculaire ou une ostéomyélite [7]. Sa potentielle gravité en fait une maladie à déclaration obligatoire aux USA et sa haute infectiosité ainsi que sa facile dissémination l'ont fait considérer comme un agent potentiel de bioterrorisme (classe B du CDC). Actuellement, la fièvre Q est traitée par l'administration de doxycycline. Bien qu'à l'origine la fièvre Q était considérée comme une maladie vectorielle à tiques [33,34], actuellement la voie d'infection par aérosol est décrite comme la voie principale [7]. Pourtant, une grande prévalence de *C. burnetii* dans les différents arthropodes laisse présumer d'un rôle potentiel de ces derniers dans la dissémination de l'agent de la fièvre Q [35]. Actuellement, un regain d'intérêt est porté à la fièvre Q du fait de sa réémergence dans différentes zones et spécialement en Europe, avec une épidémie de plusieurs milliers de cas aux Pays-Bas [36].

Le premier génome de *C. burnetii* à été séquencé en 2003 à partir d'une souche isolée d'une tique aux Etats Unis [11]. *C. burnetii* présente un génome de petite taille (environ 2 Mpb). Son annotation a permis d'identifier 2060 gènes correspondant à 77% de la séquence. La présence de nombreux pseudo gènes montre que ce génome est en cours d'évolution réductive. Très récemment, 4 nouvelles souches ont été séquencées [37]. L'ensemble des différents plasmides observés chez *C burnetii* a également été séquencé [38]. La disponibilité de ses génomes en fait un bon candidat pour entreprendre des études à l'aide de stratégies « Post génomiques » et spécialement les puces à ADN.

Comme nous l'avons précédemment évoqué, *C. burnetii* est très résistante aux différents milieux [7]. Cette haute résistance est en partie responsable d'une possible utilisation comme arme biologique. Au regard de ces données, la caractérisation des stratégies de survie de cette bactérie pour s'adapter à son nouveau milieu est importante pour la compréhension de sa pathogénicité. Dans ce contexte, nous avons

entrepris une analyse transcriptionnelle de la réponse précoce de *C. burnetii* à un changement de température. Pour ce faire, nous avons choisi d'utiliser un modèle cellulaire *in vitro* de « cold » et « heat » shock et d'analyser le profil transcriptionnel de la bactérie en utilisant la technologie des puces à ADN.

Les récentes épidémies de fièvre Q ont relancé la question de la clonalité bactérienne [39]. En effet il est possible que ces épidémies soient liées à l'émission de clones hautement virulents. De nombreuses études de biologie moléculaire ont été effectuées lors de l'épidémie qui sévit actuellement aux Pays-Bas. Ces études ont révélé l'implication d'un seul génotype ou du moins une réduction de l'hétérogénéité [36,40]. Pour essayer de mieux comprendre la pathogénicité de *C. burnetii* nous avons entrepris une analyse de génomique comparative par puce à ADN en utilisant un échantillon représentatif des souches disponibles au sein du laboratoire. Cette collection d'isolats contient notamment deux isolats présentant un génotype identique au potentiel clone épidémique des Pays-Bas ainsi que quatre souches provenant d'arthropodes.

Enfin, au cours de cette thèse, l'expertise acquise par l'utilisation des puces à ADN pour les bactéries intracellulaire nous a permis d'effectuer d'autres études sur des bactéries intracellulaires strictes pathogènes pour l'homme, les Rickettsies. Différents travaux sont présentés en annexe, avec pour point commun l'utilisation d'outils post génomique mais également la transcriptomique.

Revue

Review of Microarray Studies for Host-Intracellular Pathogens Interactions

Quentin LEROY, Didier RAOULT

Journal of Microbiological Methods

Volume 81, Issue 2, May 2010, Pages 81-95

Unité des Rickettsies, IRD-CNRS UMR 6236, Faculté de Médecine,

Marseille, France

Journal of Microbiological Methods 81 (2010) 81–95

Contents lists available at ScienceDirect

Journal of Microbiological Methods

journal homepage: www.elsevier.com/locate/jmicmeth

Review

Review of microarray studies for host–intracellular pathogen interactions

Quentin Leroy, Didier Raoult *

Université de la Méditerranée, URMITE IRD-CNRS 6236, Faculté de Médecine, 27 Bd Jean Moulin, 13385 Marseille cedex 05, France

ARTICLE INFO

Article history:
Received 23 November 2009
Received in revised form 12 February 2010
Accepted 16 February 2010
Available online 24 February 2010

Keywords:
Microarray
Pathogenicity
Virulence
Host–pathogen interactions
Obligate intracellular bacteria

ABSTRACT

Obligate intracellular bacteria are privileged soldiers on the battlefield that represent host–pathogen interactions. Microarrays are a powerful technology that can increase our knowledge about how bacteria respond to and interact with their hosts. This review summarizes the limitations inherent to host–pathogen interaction studies and essential strategies to improve microarray investigations of intracellular bacteria. We have compiled the comparative genomic and gene expression analyses of obligate intracellular bacteria currently available from microarrays. In this review we explore ways in which microarrays can be used to identify polymorphisms in different obligate intracellular bacteria such as *Coxiella burnetii*, *Chlamydia trachomatis*, *Ehrlichia chaffeensis*, *Rickettsia prowazekii* and *Tropheryma whipplei*. These microarray studies reveal that, while genomic content is highly conserved in obligate intracellular bacteria, genetic polymorphisms can potentially occur to increase bacterial pathogenesis. Additionally, changes in the gene expression of *C. trachomatis* throughout its life cycle, as well as changes in the gene expression profile of the pathogens *R. prowazekii*, *Rickettsia rickettsii*, *Rickettsia typhi*, *T. whipplei* and *C. trachomatis* following environmental changes, are discussed. Finally, an *in vivo* model of *Rickettsia conorii* within the skin is discussed. The gene expression analyses highlight the capacity of obligate intracellular bacteria to adapt to environmental changes and potentially to thwart the host response.

Contents

1. Introduction

Bacteria can be classified according to their association with their host. Obligate intracellular bacteria are unable to grow outside of host eukaryotic cells. While associations can be advantageous to their host (mutualism), other intracellular bacteria can negatively affect the cell

* Corresponding author. Tel.: +33 491324375; fax: +33 491387772.
E-mail address: didier.raoult@gmail.com (D. Raoult).

0167-7012/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.mimet.2010.02.008

(parasitism) (Tamas et al., 2001). Among intracellular bacteria pathogenic for humans, some, such as *Rickettsia* species, multiply in the cytoplasm and occasionally in the nucleus (Raoult and Roux, 1997), and some replicate in phagosomes and phagolysosomes, such as *Chlamydia* spp., *Ehrlichia* spp., *Anaplasma* spp. and *Coxiella burnetii* (Hackstadt, 1998). Understanding gene regulation and polymorphisms in these bacteria will increase our knowledge of host–pathogen interactions. Bacteria mutagenesis has been used to understand how bacteria and their hosts interact during the course of an infection. This technology can be applied to identifying and characterizing virulence-associated genes (Saenz and Dehio, 2005). However, mutagenesis studies have been historically difficult to implement in intracellular bacteria (Wood and Azad, 2000), and only recently have mutants been generated for *Chlamydia trachomatis* (Binet and Maurelli, 2009) and *C. burnetii* (Beare et al., 2009).

Genomic sequencing of intracellular bacteria (Seshadri et al., 2003; Andersson et al., 1998) and the development of robotic technologies now provide new methods to investigate the interactions between pathogens and host cells (Fournier et al., 2007). While classic molecular biology experiments focus on selected genes (Gaywee et al., 2002; Rovery et al., 2005b), microarrays provide detailed knowledge of host–pathogen interactions by high-throughput whole-genome analyses. Microarrays allow the analysis of bacterial genomic content, and the identification of genes involved in host–pathogen interactions (Cummings and Relman, 2000). Microarrays can also be used to highlight cross gene regulation between bacteria and their host during the course of an infection (Cummings and Relman, 2000; Ehrenreich, 2006).

In this review, we describe the principles of microarrays, examine their limitations as methods for studying the host–pathogen interactions of obligate intracellular bacteria and identify how the genomic diversity and gene expression of the bacteria are supportive analyses that could help to determine the pathogenicity of intracellular bacteria using genomic DNA (gDNA) or RNA hybridization microarrays.

2. Microarray technology

Massive genome sequencing has facilitated the development of many high-throughput technologies that constitute the field of functional genomics (Fournier et al., 2007). One of these, called microarray or DNA

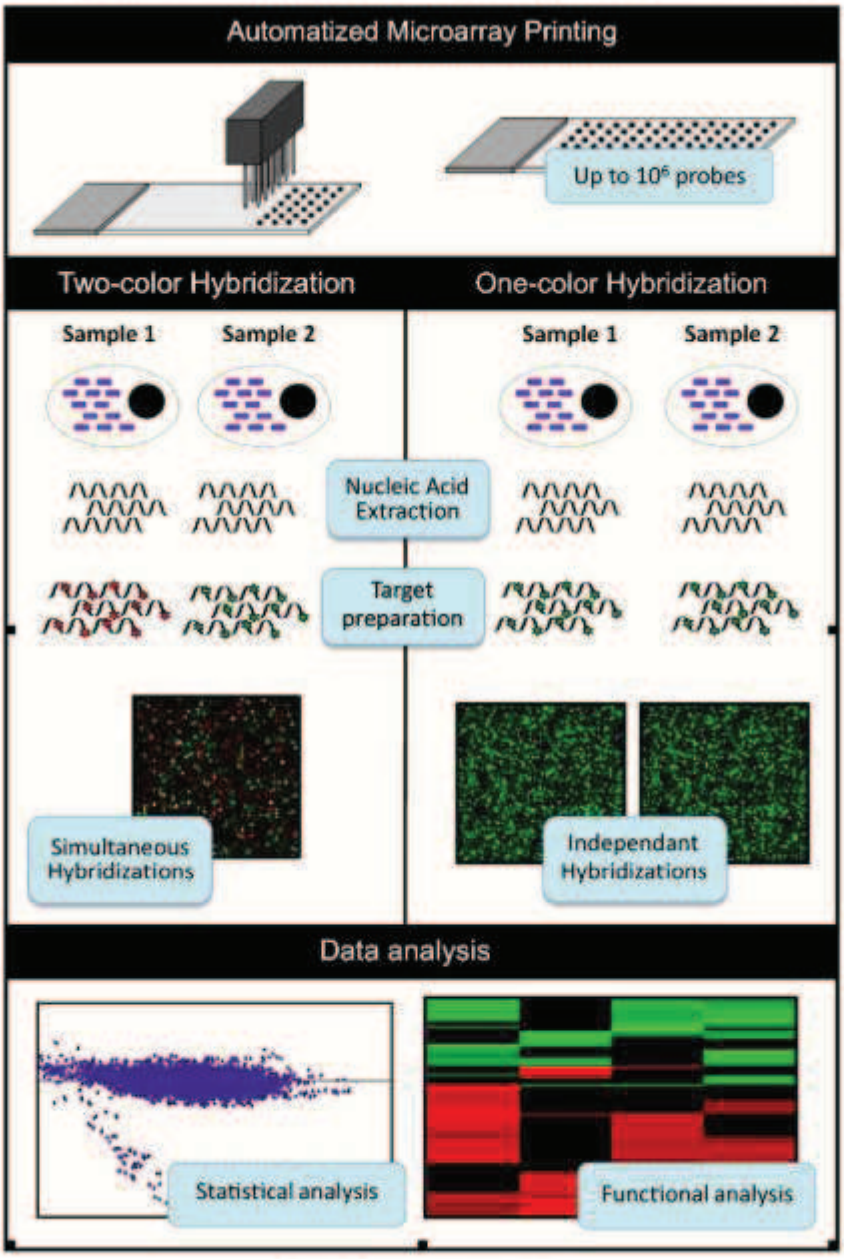


Fig. 1. Microarray experiment.

chip technology, has been used for many applications such as comparative genomic DNA hybridization (Pinkel et al., 1998), genome re-sequencing with the tiling array (Hacia, 1999) or ChIP-on-chip to identify interactions between proteins and genomic DNA (gDNA) (Ren et al., 2000). However, the most common application of this technology is gene expression profiling (Schena et al., 1995). Similar to northern or Southern blotting, the basis of a microarray is hybridization between complementary nucleic acids. In a northern or Southern blot, a labeled probe is hybridized to a nucleic acid sample immobilized on a membrane (Alwine et al., 1977; Southern, 1992). The amount of labeling from the probe that binds the complementary nucleic acid is used to compare the relative abundance of target nucleic acids. In contrast to blotting, in a microarray, the targets are labeled and the array contains a large number of probes (up to 10^6) that are attached to a solid support.

The most common microarray format is a collection of DNA samples covalently attached to a stable substrate such as a glass slide. Spots of DNA (printed DNA sequences) each represent a specific gene (or part of gene) and are arranged in an orderly pattern across the solid surface by robotic technologies (Schena et al., 1995). In gene expression profiling, the targets are comprised of a population of complementary DNA (cDNA) or complementary RNA (cRNA) sequences that are copies of the messenger RNA (mRNA), labeled with a fluorochrome. For gDNA hybridization the targets are comprised of fragmented gDNA that are labeled with a fluorochrome. The targets are applied directly to the microarray and hybridized overnight. After the hybridization step, the microarrays are scanned and images are generated. The signal intensity is proportional to the relative abundance of the target present in the starting nucleic acid sample and is quantified using a high-resolution scanner. Statistical and functional analyses are performed on the generated data (Fig. 1).

Generally microarrays are hybridized in two different ways: one-color and two-color hybridizations (Jaluria et al., 2007). Multiplexing of more than two fluorochromes has given marginal results because of emission spectra superposition for the different available fluorochromes. Two-color arrays are hybridized simultaneously with two different samples, each labeled with a different fluorochrome (Shalon et al., 1996). Generally cyanine (Cy3 or Cy5)-dCTP is incorporated in the two-color system in comparative genomic hybridization (CGH) and transcriptional analysis in microbiology (Ehrenreich, 2006). The most common experimental design for this competitive hybridization is called a control versus experiment design. In contrast, in the one-color system, the samples are hybridized separately on a single array. While one-color hybridizations decrease the experiment cost, two-color hybridizations reduce experimental variability and provide better reproducibility.

DNA microarrays are made either by printing pre-made DNA probes onto a solid surface or by synthesizing DNA probes. Improvements in printing technologies have allowed the robotic arrayers to spot approximately $1–5 \times 10^4$ fragments of PCR products or pre-synthesized oligonucleotides onto a slide (Fig. 2A). Improvements in the technology and chemistry have also allowed manufacturers to increase the probe density on on-chip synthesized oligonucleotide microarrays using techniques such as photolithography (Lipshutz et al., 1999) and ink-jet deposition (Hughes et al., 2001). These technologies produce microarrays with as many as 10^6 probes on a slide (Fig. 2B). Whereas a PCR-product microarray has a high number of false positives because of cross hybridization, oligonucleotides allow for increased specificity of hybridization when the probes have been designed appropriately. Spot density of printed array limits the number of target sequence. On chip synthesized oligonucleotide arrays that increase significantly the spot density (Fig. 2), can accommodate the entire sequence of genome or transcripts for several microbes (Selinger et al., 2000; Toledo-Arana et al., 2009).

Analysis and interpretation of microarray could be considered as a critical step. However microarray is a powerful tool, its analysis is subjected to reproducibility problem and integration of large-scale data. To compare data obtained from different microarray of a same experimental analysis, it is necessary to normalize data. Several methods have been improved to normalize efficiently microarray data (Yang et al., 2002). After data normalization statistical analyses allow to analyze data with good confidence. As normalization, statistical tests have been improved for microarray data analysis for genomic (Picard et al., 2005) or transcriptomic analysis (Tusher et al., 2001; Thomas et al., 2001). Data mining are completed with functional (Jenssen et al., 2001) and clustering analysis (Eisen et al., 1998) in order to decipher the genomic or transcriptomic variation. Software suites or independent modules are available to perform all the whole data analysis (Saeed et al., 2003; Dudoit et al., 2003).

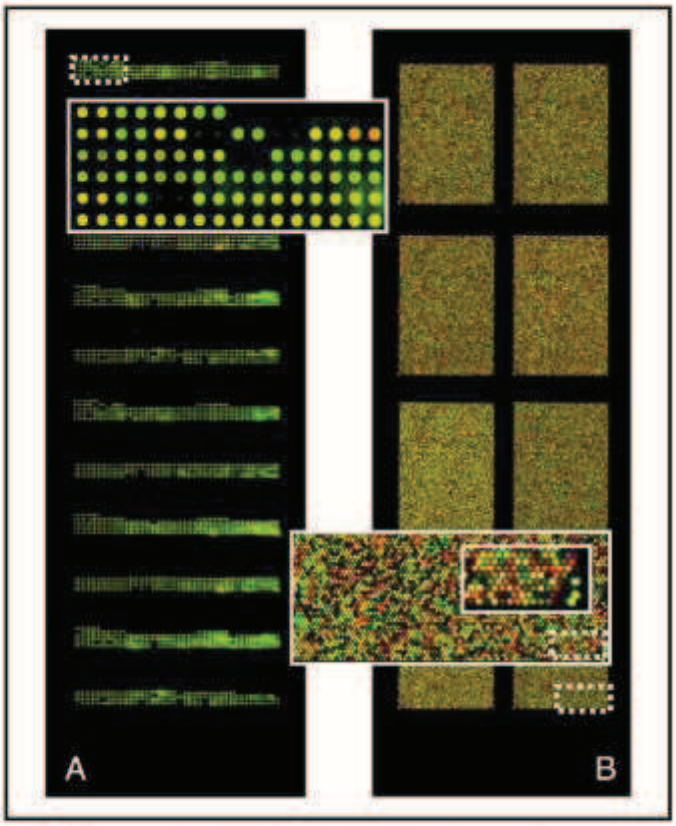


Fig. 2. Microarray density. (A) The picture represents a low-density PCR-product microarray with probes spotted by a robotic technology. (B) The picture represents a high-density oligonucleotide microarray with probes synthesized on slide by the ink-jet technology.

3. Limitations of microarrays in the study of obligate intracellular bacteria

Technical limitations are inherent when using microarrays to study host–pathogen interactions, especially in the context of obligate intracellular bacteria. Culture of intracellular is difficult. Bacteria have to be released from the cells. The amount of nucleic acids is often too low. Bacterial RNA is very labile and should be manipulated with specific cares. Moreover, for transcriptional studies, bacteria are highly sensitive to environmental changes and their transcriptomic profiles could be affected by bacteria manipulations. In this review we will describe these limitations and the solutions that have been already used in the microarray-based studies for the obligate intracellular bacteria. In this review, given the limitations and the solutions that we described afterwards, we propose a workflow to generate target for comparative genomic and transcriptional studies in Fig. 3.

As we mentioned previously, obligate intracellular bacteria are fastidious, and cell culture is required for bacterial growth. Compared to free-living bacteria or facultative intracellular bacteria such as

Escherichia coli, *Shigella* spp., *Salmonella* spp. and *Yersinia* spp., which grow in 2 or 3 days, most obligate intracellular bacteria require more than 7 days before obvious growth is detected (Merhej et al., 2009). Axenic medium has been developed for two intracellular bacteria, *C. burnetii* (Omsland et al., 2008; Omsland et al., 2009) and *T. whipplei* (Renesto et al., 2003), that now allow the growth of these bacteria free from potential contamination by eukaryotic nucleic acids.

The microarray-based DNA–gDNA hybridizations performed in comparative genomic hybridizations require purified bacterial DNA. DNA is relatively stable and does not generally require special care during purification. To release the bacteria before purification requires host cell lyses. Several procedures, including mechanical methods using sonication (Beare et al., 2006; Brunelle et al., 2004; Carlson et al., 2004) or enzymatic methods such as trypsin (Ge et al., 2004; Ge et al., 2003), can be used to lyse the host cells. Bacteria are often purified by centrifugation over a density gradient of Percol, sucrose or renographin. Micromanipulation, which allows for the dissection of vacuoles, has also been used for *C. burnetii* CGH analysis (Beare et al., 2006). Individual vacuoles were extracted from an infected monolayer. After DNA extraction, the amount of nucleic acid can be low. A whole-genomic amplification using the highly processive φ29 polymerase (GE Healthcare) can be used to overcome the low amount of bacterial gDNA (Beare et al., 2006). The different methods and technologies used in comparative genomic studies are summarized in Table 1.

Gene expression hybridizations have further limitations. RNA molecules are labile and can be rapidly degraded, resulting in a reduction or loss of many transcripts. This is particularly important for bacterial mRNA, as these nucleic acids usually have a very short half life, often only a few minutes (Condon, 2007); for example, the half life of *R. prowazekii* RNA is estimated to be about 15 min (Winkler, 1987). As soon as RNA is extracted it needs to be stored at a low temperature (−20 °C or −80 °C), but be manipulated on ice with special care. Different storage buffers prevent RNA degradation, including 0.1 mM EDTA or Tris–EDTA buffer (TE) and commercial buffers such as THE RNA Storage Solution (Ambion). Importantly, transcriptional profiles can be modified during the handling and processing of bacteria. Therefore, the samples need to be manipulated and stored with care. Harvested or purified bacteria need to be frozen in liquid nitrogen to decrease the incidence of RNA profile changes. Commercial buffers have been improved to overcome RNA degradation and transcriptional alterations during storage and sample manipulation. RNAprotect Bacteria Reagent included in the RNeasy Protect Bacteria Kit (Qiagen) provides transcriptional stability during the handling of bacteria. Additionally, RNAlater (Ambion) can be used for tissue storage for many months; infected skin biopsies have been used to perform a whole-transcriptional analysis of an *R. conorii* infection after being stored in RNAlater for 3 years (Renesto et al., 2008).

The critical problem in the analysis of obligate intracellular bacteria is that both nucleic acids of the host cell and pathogen are present in the same sample. The different proportions of eukaryotic and prokaryotic transcripts that can occur in the course of an infection are a limiting factor for the hybridization efficiency in the analysis of intracellular bacteria (Fig. 4A). Because eukaryotic mRNA competes with bacterial mRNA during the labeling step, microarray technology requires purified samples (La et al., 2008; Hinton et al., 2004). It has been shown for *R. conorii* that the high proportion of host cell transcripts (up to 90%) results in the exhaustion of fluorochromes. Eukaryotic contaminations result in a low efficiency in the labeling of

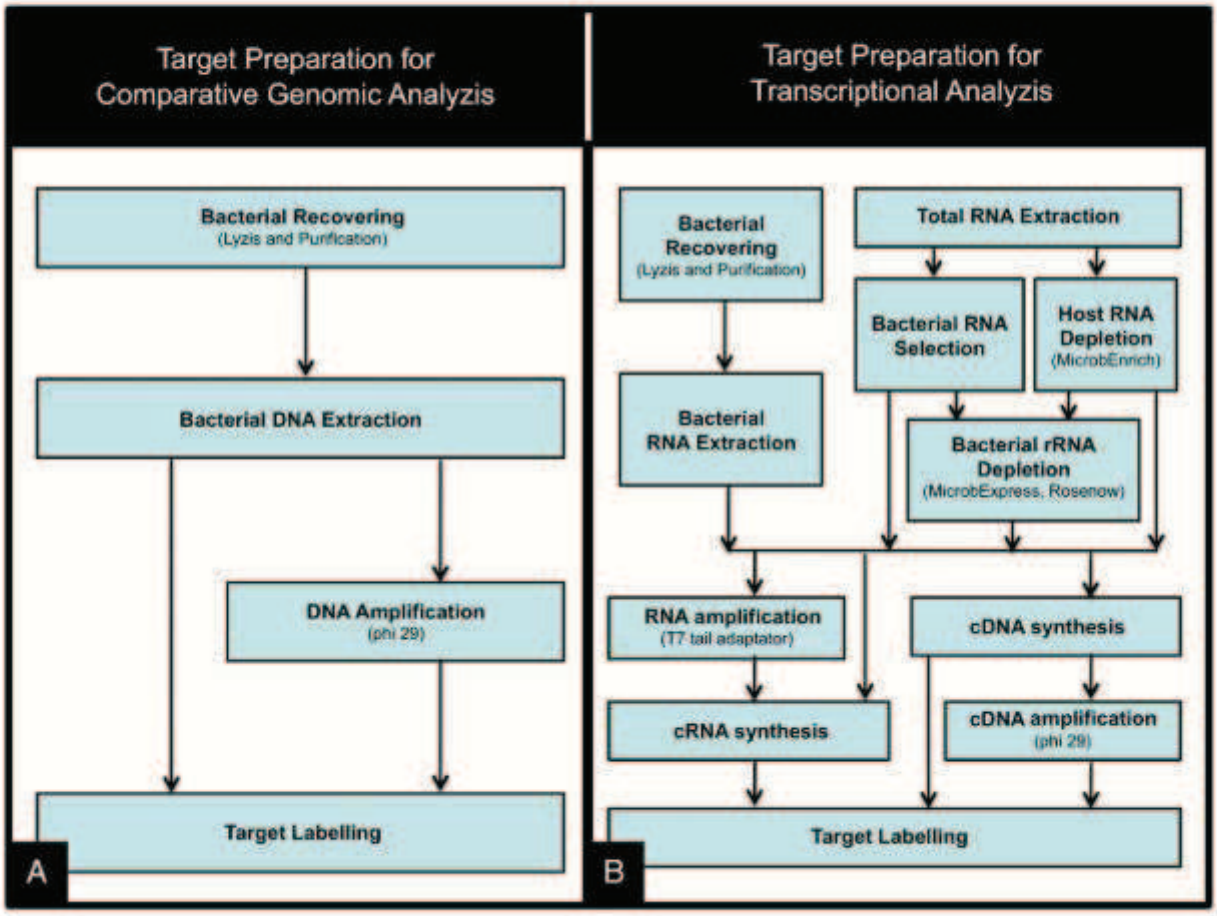


Fig. 3. Workflow to perform microarray experiments in host–pathogen interaction field. (A) The picture represents the strategies to perform comparative genomic analysis. (B) The picture represents the strategies to perform gene expression analysis.

Table 1
Strategies used in microarray hybridizations with intracellular bacteria.

Bacteria	Studies	Lysis	Purification	Labelling	Amplification	Microarray	References
Chamydia trachomatis	GC	Sonication	Renographin density gradient	Two colors with Cy3/5 dCTP	–	PCR-product spotted array	Brunelle et al. (2004)
	GC	Sonication	Renographin density gradient	Two colors with Cy3/5 dCTP	–	Oligonucleotide spotted array	Carlson et al. (2004)
	GE	Sonication	Renographin density gradient	Two colors with Cy3/5 dCTP	–	PCR-product spotted array	Nicholson et al. (2003) and Nicholson et al. (2004)
	GE	Sonication	Renographin density gradient	Two colors with Cy3/5 dCTP	Specific primer amplification	Oligonucleotide spotted array	Belland et al. (2003a) and Belland et al. (2003b)
Coxiella burnetii	GC	Sonication	Sucrose and renographin density gradient/microdissection	One color with biotin-ddUTP	phi29 amplification	In situ photolytographic synthesis oligonucleotide array	Beare et al. (2006)
	GE	Trizol	MicrobEnrich/axenic medium	One color with biotin-ddUTP	polyA tails ligation and polydT amplification	In situ photolytographic synthesis oligonucleotide array	Omsland et al. (2009)
Ehrlichia chaffeensis	GC	Nitrogen cavitation	Percoll density gradient	Two colors with Cy3/5 dCTP	–	Oligonucleotide spotted array	Miura and Rikihisa (2007)
Mycobacterium leprae	GE	Trypsination and Zirconia beads	Differential centrifugation	One color with fluorescent 9mer	–	In situ maskless light projector synthesis oligonucleotide array	Akama et al. (2009)
Rickettsia conorii	GE	RLT and Lyzozyme	MicrobEnrich	Two colors with Cy3/5 dCTP	phi29 amplification	PCR-product spotted array	La et al. (2007b)
	GE	RLT and proteinase K	MicrobEnrich	Two colors with Cy3/5 dCTP	phi29 amplification	In situ ink-jet synthesis olignucleotide array	Renesto et al. (2008)
Rickettsia prowazekii	GC	Trypsination	Renographin density gradient	Two colors with Cy3/5 dCTP	–	PCR-product spotted array	Ge et al. (2003) and Ge et al. (2004)
	GE	Trypsination	MicrobEnrich	Two colors with fluorescent dendrimers	–	Oligonucleotide spotted array	Audia et al. (2008)
Rickettsia rickettsii	GE	Trizol	–	One color with biotin-ddUTP	–	In situ photolytographic synthesis oligonucleotide array	Ellison et al. (2009)
Rickettsia typhi	GE	Sonication	Renographin density gradient	Two colors with Cy3/5 dCTP	–	In situ photolytographic synthesis oligonucleotide array	Dreher-Lesnick et al. (2008)
Tropheryma whipplei	GC	Axenic medium	Axenic medium	Two colors with Cy3/5 dCTP	–	PCR-product spotted array	La et al. (2007a)
	GE	Axenic medium	Axenic medium	Two colors with Cy3/5 dCTP	–	PCR-product spotted array	Crapoulet et al. (2006)
	GE	Axenic medium	Axenic medium	Two colors with Cy3/5 dCTP	–	PCR-product spotted array	Van et al. (2007)

GC: genomic comparison.
GE: gene expression.

bacterial transcripts, and preclude microarray hybridization analysis (La et al., 2007b). Therefore, host nucleic acid removal is required to successfully perform bacterial hybridizations.

To remove eukaryotic RNA contamination, microorganisms can be purified before RNA extraction after liberation from infected host cells. Several lysis procedures may be used to disrupt the host cells. Mechanical lysis using cell disruption by sonication can be used (Nicholson et al., 2003). Alternatively, the cell membrane can also be lysed using chemicals such as saponin (Grifantini et al., 2002), guanidine thiocyanate (Schnappinger et al., 2003), sodium dodecyl sulfate (SDS) coupled with a phenol/ethanol mixture (Eriksson et al., 2003), RLT lysis buffer (Qiagen) (Di et al., 2005) or trypsin–EDTA (47). The purification of bacteria from infected organs, such as the lung (Talaat et al., 2007), requires tissue homogenization. Once released, bacteria are separated from cell debris by differential centrifugation over a density gradient such as sucrose or renografin. These procedures are often long and difficult and can affect the original transcriptional profile. Notably, the processing is often performed at 4 °C or on ice, and bacteria are highly sensitive to cold shock and temperature stress (Weber et al., 2005).

The immediate simultaneous extraction of both pathogen and host cell mRNA ensures stability. Usually this strategy is used to monitor the transcriptional profile of a few genes by RT-PCR (Rovery et al., 2005b; Rovery et al., 2005a). Therefore, while eukaryotic contaminant RNA does not alter the results obtained using RT-PCR, these contaminants have to be removed when performing microarray-based experiments to circumvent the fluorochrome exhaustion mentioned previously. Different strategies have been developed to overcome the eukaryotic RNA contamination problem. The depletion of contaminant eukaryotic RNA has been achieved mostly through the use of subtractive hybridization technology developed by Ambion (MicrobEnrich Kit). This approach is based on capturing oligonucleotides coupled with magnetic beads that hybridize to the 18S and 28S ribosomal RNA and the poly-adenylated 3′ tail of eukaryotic mRNA. It has been applied successfully to purify prokaryotic RNA from infected eukaryotic cells (La et al., 2007b; Garzoni et al., 2007), animal tissues (Tuanyok et al., 2006) and clinical samples (Renesto et al., 2008), allowing for the microarray-based analysis of specific bacterial transcription profiles *in vivo*. Examples of eukaryotic RNA depletion are shown in Fig. 4. Because the amount of clinical sample is often limited, it may be impossible to divide the sample to analyze simultaneously the host and pathogen transcriptional profiles. The recovery of eukaryotic RNA is possible with the hybridization capture developed by Ambion. It has been suggested that host and pathogen transcriptional studies from the same RNA sample extracted from skin biopsies infected with *R. conorii* are possible (Leroy et al., 2009).

The proportion of bacterial nucleic acids may vary among pathogens and according to the degree of infection, but the amount of purified prokaryotic nucleic acid is often low. Another challenge in studying the host–pathogen interactions using DNA microarrays is obtaining a sufficient amount of nucleic acids to perform the hybridizations. In the case of gene expression, while the amplification of mRNA from eukaryotic cells using T7 polymerase amplification methods has improved (Van Gelder et al., 1990) (Wang et al., 2000), different approaches exist to increase the proportion of prokaryotic mRNA or to amplify bacterial RNA (Francois et al., 2007). With partial depletion of rRNA, the proportion of RNA that competes with mRNA during the labeling step will decrease. A method known as MicrobExpress (Ambion), based on the same principle as MicrobEnrich, allows for the enrichment of the proportion of bacterial mRNA by capturing ribosomal and transfer RNA. The enrichment of bacterial mRNA can also be achieved by selective degradation of rRNA (Rosenow et al., 2001). By synthesizing first-strand cDNA selectively from rRNA with specific primers, RNAse H digestion and then DNAse I digestion remove successively rRNA and cDNA, resulting in an enrichment of mRNA. The amplification of bacterial RNA can be achieved through enzymatic incorporation of poly A tails to the RNA molecules (Ambion), followed by linear amplification using T7 polymerase (Lathem et al., 2005; Motley et al., 2004). Another strategy to overcome the limitation of a low amount of starting material is based on the selective amplification of bacterial cDNA using genome directed primers designed with a dedicated computer algorithm (Talaat et al., 2000). Unbiased amplification of purified prokaryotic RNA has also been reported using conserved and specific primers (Belland et al., 2003a;Belland et al., 2003b), as well as using random priming coupled with the φ29 phage polymerase (La et al., 2007b; Francois et al., 2007). The different methods and technologies used in transcriptional studies are summarized in Table 1.

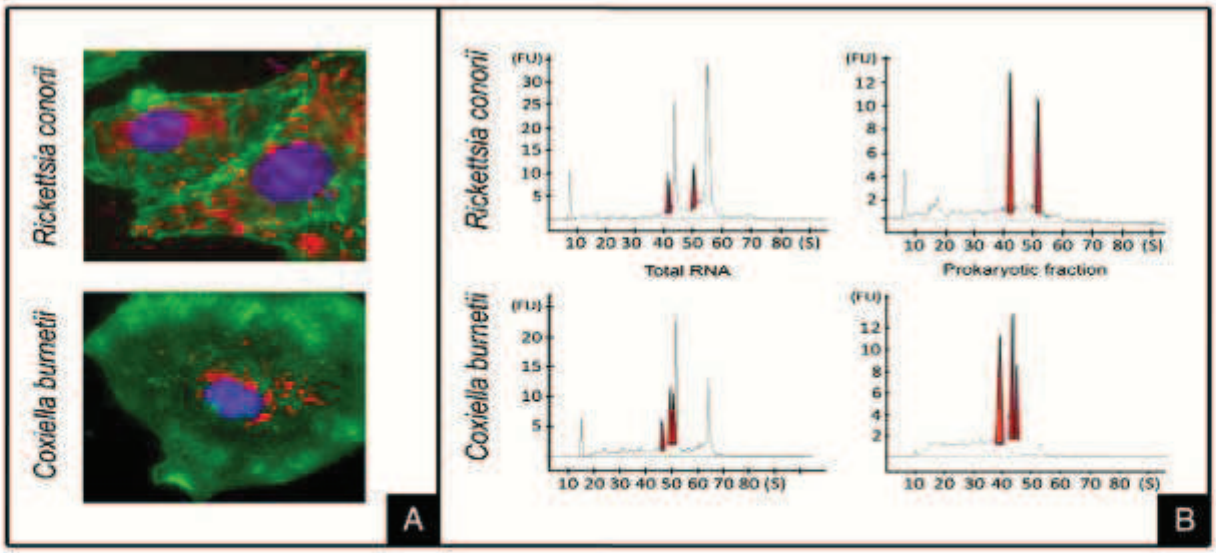


Fig. 4. Depletion of eukaryotic contaminant RNA. (A) The picture represents IF of *R. conorii* infecting Vero cells and illustrating the low proportion of bacterial nucleic acids compared to eukaryotic nucleic acids. Blue: nucleic acid stained with DAPI; red: rickettsial organisms were labeled with rabit anti *R. conorii* polyclonal antibodies; green: filamentous actin was labeled with FITC-phalloidin. (B) The figure represents the RNA quality electrophoregrams of *R. conorii* or *Coxiella burnetii*. The electrophoregrams, before and after MicrobEnrich treatment, allow to see the depletion of eukaryotic RNA. The bacterial ribosomal RNA was painted in red.

Table 2
Obligate intracellular bacteria.

Bacteria	Disease	Host cell type	Genome sequence	Genome size	ORF number	References
Chlamydia trachomatis	Trachoma, urethritis and lymphogranuloma venereum	Epithelial cell	11	1 Mbp	893	Storz and Spears (1977) and Stephens et al. (1998)
Coxiella burnetii	Q fever	Macrophage and monocyte	5	2 Mbp	2092	Maurin and Raoult (1999) and Seshadri et al. (2003)
Ehrlichia chaffeensis	Human monocytic ehrlichiosis	Macrophage, monocyte and leukocyte	2	1.18 Mbp	1115	Paddock and Childs (2003) and Hughes et al. (2001)
Mycobacterium leprae	Leprosy	Schwann cells, macrophages, and other host cells	1	3.3 Mbp	1605	Walker and Lockwood (2007) and Colston (2001)
Rickettsia conorii	Mediterranean spotted fever	Endothelial cell	1	1.26 Mbp	1374	Rovery and Raoult (2008) and Ogata et al. (2001)
Rickettsia prowazekii	Epidemic typhus	Endothelial cell	2	1.11 Mbp	834	Bechah et al. (2008) and Andersson et al. (1998)
Rickettsia rickettsii	Rocky Mountain spotted fever	Endothelial cell	2	1.26 Mbp	1345	Dantas-Torres (2007) and Ellison et al. (2008)
Rickettsia typhi	Murine typhus	Endothelial cell	1	1.11 Mbp	877	Civen and Ngo (2008) and McLeod et al. (2004)
Tropheryma whipplei	Whipple's disease	Macrophage and monocyte	2	0.93 Mbp	808	Fenollar et al. (2007) and Bentley et al. (2003)

4. Microarray analysis

4.1. Comparative genomic analysis

CGH experiment design for obligate intracellular bacteria may be split into two different parts. The first concerns the comparison within the same species of virulent and avirulent strains. The second one concerns a strain collection screening within the same species, causing either different diseases (*C. trachomatis*) or different clinical symptoms (*T. whipplei* and *C. burnetii*). Finally, we will highlight the principal genomic variations that can be observed and potentially associated with host–pathogen interactions with obligate intracellular bacteria.

4.1.1. Virulence comparison

Epidemic typhus is a louse-borne disease caused by *Rickettsia prowazekii* (Table 2). *Pediculus humanus corporis*, the body louse, is the main known vector for epidemic typhus (Bechah et al., 2008). Epidemic typhus has been considered a disease of the past, but outbreaks and sporadic cases can occur when sanitary conditions are conducive to lice proliferation such as war, famine and social disruption. Epidemic typhus is characterized by a rash, fever and headaches and does not occur directly by bites of the body louse but mainly by contamination of bite sites with the feces or crushed bodies of infected lice. *R. prowazekii* is a Gram-negative intracellular alpha-proteobacterium, belongs to the typhus group from the *Rickettsia* genus, multiplies in endothelial cells of vessels, lacks intracellular motility and multiplies freely in the host cell cytoplasm (Raoult and Roux, 1997). The 1.11 Mbp *R. prowazekii* (Madrid E strain) genome was sequenced in 2001 and codes for 834 ORFs that represent 76% of the coding content. Like other intracellular bacteria, *R. prowazekii* has a small genome and an evolutionary tendency towards genomic reduction (Andersson et al., 1998). A comparative genomic approach was undertaken to highlight the virulence of *R. prowazekii* using a whole genome microarray of PCR products (Ge et al., 2004; Ge et al., 2003). Two different strains, BreinL and Madrid E, that are a virulent and a vaccine strain, respectively, were compared. The comparative analysis showed that the genomes of the avirulent and virulent strains were highly conserved. Only 3% of the genome was different between the two strains. One deletion of the *rp084* gene was observed in the virulent strain. This gene has no known function, but an analysis of the putative structure of the protein encoded by *rp084* suggests that it has homologies with proteins implicated in cytochrome-*c*. Cytochrome-*c* is an important factor in the apoptotic cascade. Several genes where found to be different between the two strains because of the presence of paralogs, especially for known virulence-associated genes in other bacteria. The gene duplications of virulence-associated genes and the loss of a specific gene seem to increase the virulence of *R. prowazekii*. This comparative genomic analysis demonstrates how a few differences in genomic content can significantly affect virulence.

Human monocytotropic ehrlichiosis (HME) is a tick-borne disease caused by *Ehrlichia chaffeensis* (Table 2). *Amblyomma americanum*, the lone star tick, is the main known vector for *E. chaffeensis* (Anderson et al., 1993). Wild mammals are the main reservoir of *E. chaffeensis*. HME has been reported in many parts of the globe. HME is characterized by fever, headache, myalgia, anorexia and chills (Paddock and Childs, 2003). *E. chaffeensis* is a Gram-negative intracellular alpha-proteobacterium that multiplies as micro-colonies in vacuoles within monocytes, macrophages and polymorphonuclear leukocytes (Walker et al., 2004). The 1,18 Mbp. *E. chaffeensis* genome was sequenced in 2006 and codes for 1115 ORFs that represent 79.7% of the coding content. As an intracellular bacterium, *E. chaffeensis* has a small genome and an evolutionary tendency towards genomic reduction (Hughes et al., 2001). A comparative genomic approach was undertaken to highlight the virulence of *E. chaffeensis* using a whole genome short oligonucleotide tiling microarray (Miura and Rikihisa, 2007). Two different strains

(Wakulla and Liberty) were compared to the Arkansas strain. The pathogenicity of these strains, which represent the three genomic groups of *E. chaffeensis*, was explored in parallel, and the Arkansas, Wakulla and Liberty strains presented with low, high and medium virulence, respectively. The comparative genomic analysis showed that the three strains were highly conserved. The genomes of the Wakulla and Liberty strains contained a 30 kbp region that was highly divergent. This region contains a cluster of genes coding for major outer membrane proteins. Additionally, two ankyrin repeat proteins were also divergent among the three strains. The *E. chaffeensis* virulence phenotype seems to be correlated with its ability to communicate with its host using membrane protein modifications.

4.1.2. Diversity of genomic content for collection of bacterial strains

Whipple's disease is caused by *Tropheryma whipplei* (Table 2), but a genetic predisposition has been proposed (Fenollar and Raoult, 2003; Fenollar et al., 2007). The reservoir is suspected to be environmental. Whipple disease is typically characterized by arthralgia, weight loss, diarrhea and abdominal pain (Schneider et al., 2008). *T. whipplei* is a small Gram-negative intracellular actinomycete that has been observed within intestinal macrophages and circulating monocytes (Raoult et al., 2001). The 0.93 Mbp *T. whipplei* genome was sequenced in 2003 (Bentley et al., 2003; Raoult et al., 2003) and codes for 808 ORFs that represent 86% of the coding content. The analysis of the *T. whipplei* genome has allowed for improvement of an axenic medium, making the culture and isolation of *T. whipplei* easier (Renesto et al., 2003). As an intracellular bacterium, *T. whipplei* has a small genome and an evolutionary tendency towards genomic reduction. The genetic diversity of *T. whipplei* (Table 1) was assessed through a genomic comparison using a PCR-product microarray (La et al., 2007a). Fifteen isolates originating from various geographic and biological sources were compared to the Twist isolate. The genomic comparison of the different strains revealed only limited genomic variation, with up to 2.2% of the ORFs found to be polymorphic, in contrast to enteric bacteria, which have 15% to 23% variation. Thirty-four genes were lacking or divergent, and most have an unknown function. The genomic comparison showed that the WISP proteins, which are heterogeneous surface proteins containing WND repeats, are highly divergent across the *T. whipplei* strains. Variations in the WISP proteins suggest that they could be responsible for various pathogen features and evasion of the host immune system. A large deletion has been found in one *T. whipplei* strain in a genomic inversion previously described (Raoult et al., 2003). Genes encoding WISP proteins flank the genomic fragment, highlighting a putative role for the WND repeats in *T. whipplei* genome plasticity. This analysis highlights the mechanisms of *T. whipplei* adaptation to their environment. Variations in the WISP protein family could play a key role in virulence. The described deletion confirms the reductive evolution of *T. whipplei*.

Q fever is a zoonotic disease that is distributed worldwide and caused by *C. burnetii* (Table 2). The *C. burnetii* reservoir is large and includes many wild and domestic animals and arthropods. The main infection route for human infection is the inhalation of aerosolized parturient fluids from infected animals. Q fever can manifest as an acute or chronic illness. Acute Q fever is typically a self-limiting febrile illness during which pneumonia or hepatitis can occur, whereas chronic Q fever is a severe illness that can present endocarditis, vascular infection, osteomyelitis and chronic hepatitis (Maurin and Raoult, 1999; Raoult et al., 2005). *C. burnetii* is a Gram-negative intracellular gamma-proteobacterium that multiplies within acidic vacuoles of eukaryotic cells. *C. burnetii* displays antigenic variation in lipopolysaccharide (LPS). Phase I LPS is highly infectious and corresponds to the natural phase found in animals, humans and arthropods, whereas phase II LPS is not very infectious and can be obtained after several passages through cell culture or fertilized eggs. The approximately 2 Mbp *C. burnetii* genome was sequenced in 2003 and codes for 2092 ORFs that represent 89.1% of the coding content, while its plasmid codes for 40 ORFs (Seshadri et al., 2003). As with other intracellular bacteria, *C. burnetii* has a small genome that shows an evolutionary tendency towards genomic reduction. The genetic diversity of *C. burnetii* (Table 1) was assessed in a genomic comparison using a short oligonucleotide array (Beare et al., 2006); 24 isolates from diverse sources such as humans, animals or arthropods representing acute, endocarditis, abortion or hepatitis diseases were compared in the study to the Nine Mile phase I reference isolate (NMI). The NMI is biologically representative of human acute disease isolates. The CGH analysis revealed that the *C. burnetii* genomic content is highly conserved. Among all *C. burnetii* isolates, 139 chromosomal and plasmid ORFs were polymorphic, representing 7% of the NMI coding content. The degree of variance between NMI and chronic disease-related isolates ranged from 2.1% to 4.1%. The deleted ORFs are mostly hypothetical or annotated as non-functional. Five missing ORFs in several chronic isolates were predicted to encode proteins with ankyrin or tetratricopeptide repeats. These proteins are suspected to play a key role in host–pathogen interactions. Two ORFs encoding proteins that have homology with the quorum-sensing regulator LuxR were deleted in the chronic isolates. A few polymorphic ORFs associated with chronic disease could result in slower growth and, consequently, an attenuated immune response. The large deletion implicated in the LPS variation previously described (Hoover et al., 2002) is not required for the LPS phase variation. While *C. burnetii* seems to be able to adapt to its environment by short genomic variations in targeted genes, the virulence factor(s) responsible for its pathogenicity remains unclear.

C. trachomatis is a Gram-negative obligate intracellular bacterium that exhibits a tropism for conjunctival and urogenital columnar epithelial cells (Schachter, 1978) (Table 2). *C. trachomatis* replicates in vacuoles and is characterized by a complex intracellular developmental cycle that involves a metabolically inactive, non-replicating infectious form called an elementary body (EB). After entry into the target cell, the EB differentiates into a metabolically active replicating form that is called the reticulate body (RB) (Storz and Spears, 1977). *C. trachomatis* isolates can be differentiated into three different biovars that can cause trachoma, cervitis and urethritis and finally lymphogranuloma venereum (LVG) (Schachter, 1978). Eleven genomes of *C. trachomatis* have been sequenced, and the 1 Mbp *C. trachomatis* genome encodes 893 chromosomal and 8 plasmid ORFs (Stephens et al., 1998). To understand how the high tropism of *C. trachomatis* (Table 1) could be correlated to genomic variation, several CGH analyses were performed separately (Carlson et al., 2004; Brunelle et al., 2004). The genomes of 15 serovars distributed in all biovars were compared to the serovar D that is associated with urethritis. Overall, gene conservation was very high; up to ten genes have been found to be divergent or absent in one serovar. These findings suggest that the tissue tropism or virulence traits of *C. trachomatis* are controlled by a small set of genetic differences. While *ompA* is used for the serological classification of a serovariant, its genetic variability does not correlate with the tropism of a *C. trachomatis* biovar. Rather, a highly divergent region of the genome, which has been already described (Read et al., 2000), correlates with the bacterial tropism. The region contains a gene (*CT166*) that codes for an ORF with homology to toxin-like genes and the *trp* operon. To precisely define how the toxin-like gene could explain the tropism of *C. trachomatis*, the *CT166* locus was sequenced (Carlson et al., 2004). The analysis of the cytotoxin region showed that each serovar contained a unique combination of DNA sequences. The strict association of complete or partial toxin genes with the chlamydial biovars that exhibit specific tropism suggests that the toxin could be an important virulence factor responsible for the tropism.

4.1.3. Genomic variation

Comparative genomic analyses have revealed that genomic content is highly conserved among different strains of obligate intracellular bacteria. As suggested by Andersson and Ochman (Andersson and Andersson, 1999; Ochman and Moran, 2001), obligate intracellular bacteria are

generally separated from other bacteria that might provide genes and genetic exchange. This lack of interactions with other bacteria limits evolutionary events and, therefore, genomic diversity. Deletions of genomic content, found in the CGH analysis, confirm the paradigm of genomic reduction exhibited by obligate intracellular bacteria. However, microarrays can only monitor spotted sequences and cannot reveal additional genomic material except for some paralogous genes, as shown with *R. prowazekii* (Ge et al., 2004). We have observed that genomic diversity is concentrated in a specific, highly divergent region. These genomic regions are hot spots of plasticity because of the presence of repeat sequences that may increase genomic mutation. The capacity to reveal divergence depends on the type of microarray. PCR-product microarrays allow for only the identification of large genomic variations such as gene deletions; high-density oligonucleotide microarrays can identify single nucleotide polymorphisms. Microarray-based studies have allowed investigators to focus on bacterial pathogenicity. Genes that code for membrane proteins seem to be preferentially modified in intracellular bacteria (Table 3). As proposed by Lin et al. (2002), outer membrane proteins play a key role in the adaptation of bacteria to their hosts. It is likely that membrane proteins are associated with tropism, as has been proposed with the *C. trachomatis omp* family (Carlson et al., 2004) and the *T. whipplei WISP* family (La et al., 2007a). The polymorphism, in the gene coding for a cytotoxin-like protein (CT1666), is also associated with specific biovars (Carlson et al., 2004). The protein has a glycosyltransferase domain, suggesting that the protein could be involved in the synthesis of a glycoprotein or in the lipopolysaccharide variation, which is known to be important for bacterial tropism during the host–pathogen interaction. Genes encoding eukaryotic-like proteins, including ankyrin proteins, are also targets for genomic variation in different isolates of obligate intracellular bacteria (Beare et al., 2006; Miura and Rikihisa, 2007). These proteins have domains that are typically found in eukaryotic proteins and facilitate protein–protein interactions (Lin et al., 2002; Goebl and Yanagida, 1991). Recently Pan et al. (2008) have shown that ankyrin is excreted into host cells and possesses effector functions important for the bacterial infection of eukaryotic host cells. Recently, ankyrin repeats have been associated with the Dot/Icm secretion system for *C. burnetii* (Voth et al., 2009). The genes coding for proteins that are suspected to interact with the host seem to be preferentially modified or deleted, suggesting that host–pathogen interactions are essential in the virulence of these bacteria (Table 3).

4.2. Transcriptional analysis

Transcriptional analysis design for obligate intracellular bacteria may be split into 3 different parts. The first concerns the transcriptome of the bacteria in course of infection. The second one concerns a bacteria submitted to environmental changes such as temperature variation. The third part corresponds to an *in vitro* analysis of *R. conorii* from clinical samples. Finally, we will focus on the principal genetic variations that can be observed and potentially associated with host–pathogen interactions with obligate intracellular bacteria.

4.2.1. In course of infection

Several gene expression microarray-based studies have explored the global gene expression of *C. trachomatis* in the course of an infection. These transcriptomic analyses have allowed for the identification of the mechanisms involved in the complex developmental cycle of *C. trachomatis*. The identification of genes that are regulated during the *C. trachomatis* life cycle was reported in two separate studies using L929 fibroblast cells (Nicholson et al., 2003) or HeLa epithelial cells (Belland et al., 2003b). While differences were observed in these studies, three major stage-specific gene sets were apparent. During the early stage, only a small set of genes, including *groEL* and *euo*, is transcribed. The low transcriptional activity could be explained by pre-packaged proteins that are post-translationally activated and participate in early biological events following infection (Nicholson et al., 2003). The middle stage appeared to be a major transcriptional switch that correlated with the RB to EB transition. The late stage was characterized by the reorganization of membranes and the formation of a condensed nucleoid. Nicholson et al. suggested a late assembling of the Type III secretion. We hypothesize that the differences exhibited by the two transcriptional analyses could be explained not only by differences in the techniques used to perform the microarrays, but also by the tropism of the *C. trachomatis* to the host cell.

Mycobacterium leprae is the causative agent of leprosy also known as Hanson's Disease (Table 2). Patients commonly present with skin lesions, numbness or weakness caused by peripheral nerve involvement, or more rarely, a painless burn or ulcer in an anesthetic hand or foot. Patients are highly infectious, and the organism can remain viable outside a human host for many months (Walker and Lockwood, 2007). *M. leprae* is a Gram-positive bacterium that has never been grown in cellular system culture. Its target cell seems to be macrophages, dendritic and Schwann cells (Barker, 2006). *M. leprae* will grow in the footpads of mice and in armadillos (Walker and Lockwood, 2007). The culture can take several weeks to mature. The 3.3 Mbp *M. leprae* genome was sequenced in 2001 (Colston, 2001). The number and ratio of pseudogenes in *M. leprae* genome are exceptionally large with 1605 genes and 1115 pseudogenes, while M. tuberculosis has 3959 genes and only 6 pseudogenes. A transcriptomic analysis of *M. leprae* infecting nude rat was assessed using an oligonucleotide tiling array that covers the entire *M. leprae* genome (Akama et al., 2009). Tiling arrays were used to analyze comprehensive RNA expression of not only genes, but also pseudogenes and non-coding regions in *M. leprae*. In order to confirm the specificity of the tiling array, RNA from the same sample was simultaneously hybridized with the ORF array on which multiple sequence-specific probes were designed for each gene. The positive signals detected on the ORF array were consistent with those detected on the tiling array. The mean intensity in coding genes was significantly lower than that in non-coding regions and pseudogenes. High RNA expression from these non-coding regions suggests that those RNAs have a biological function.

4.2.2. Environmental stress

Belland et al. (2003a) investigated the transcriptional response of *C. trachomatis* to IFN-γ exposure, simulating persistence and reactivation from persistence. Altered but active biosynthetic processes characterized persistence especially with an increase in the *trp* operon transcription, suggesting an important role for tryptophan. Replication seems to continue. Upon removal of IFN-γ, the transcriptional changes were reversed, and *C. trachomatis* returned to its classic developmental cycle. Whereas bacterial persistence was associated with a general stress response, the authors suggest that persistence is an alternative life cycle used by *chlamydiae* to overcome the host immune response. Nicholson et al. (2004) have shown that *C. trachomatis* was not transcriptionally sensitive to carbon starvation, in contrast to free-living bacteria.

More recently, the transcriptional response of *R. prowazekii* to heat shock (30 min at 42 °C) was analyzed using a whole genome oligonucleotide microarray (Audia et al., 2008). Compared to 34 °C, only 23 genes were differentially upregulated, whereas no gene was significantly downregulated. The temperature increase induced the upregulation of proteases and chaperone proteins, suggesting that *R. prowazekii* has mechanisms to allow for optimal folding of proteins during heat shock. In addition, the upregulation of a gene encoding a DNA repair protein suggests that *R. prowazekii* DNA integrity is sensible to temperature stress. The upregulated genes that have no known role in the heat stress response are immediately downstream of heat shock genes with a lower induction. Overall, these results that *R. prowazekii* adapts to growth at 42 °C through the modulation of

Table 3
Principal results in microarray hybridizations with intracellular bacteria.

Bacteria	Studies	Experiments	Results	References
Chamydia trachomatis	GC	Comparison of 15 serovars	- Genomic content is highly conserved - Plasticity zone contains the cytotoxin like gene CT166 and the trp operon - Polymporphism in ompA and pmp genes coding for membrane protein	Brunelle et al. (2004) and Carlson et al. (2004)
	GE	Developpmental cycle in course of L929 and HeLa cell infection	- *Early stage*: few genes comprising groEL and euo - *Middle stage*: major switch corresponding to RB from EB transition - *Late stage*: reorganization to the membrane and nucleoid condensation	Nicholson et al. (2003) and Belland et al. (2003a)
	GE	Response to IFN-gamma treatment	- Bacteria persistence associated with an increase of the tryptophan operon expression	Belland et al. (2003b)
	GE	Response to carbon starvation	- No differences	Nicholson et al. (2004)
Coxiella burnetii	GC	Comparison of 24 isolates	- Genomic content is highly conserved - Several eukaryotic like proteins are highly divergent - 2 quorum sensing regulator (luxR family) deleted in chronic disease	Beare et al. (2006)
	GE	Comparison of *C. burnetii* growing in CCM medium compared to growing in Vero cells	- Downregulation of the ribosomal protein suggesting that protein synthesis is insufficient to support the *C. burnetii* replication	Omsland et al. (2009)
Ehrlichia chaffeensis	GC	Comparison of 3 strains representing high, medium and low virulence phenotypes	- Genomic content highly conserved - A 30 kb highly divergent containing membrane protein and ankyrin repeat protein	Miura and Rikihisa (2007)
Mycobacterium leprae	GE	Infection of a nude rat	- RNA expression was detected from genes, pseudo-genes, and noncoding regions	Akama et al. (2009)
Rickettsia conorii	GE	Response to carbon starvation	- [p]ppGpp accumulation	La et al. (2007b)
	GE	Bacteria from infected skin of MSF patient compared to *in vitro* growth within Vero cells	- Transcriptomic profile highly conserved among different eschars - Transcriptomic profile mainly downregulated and associated with bacteria growth arrest - Upregulated genes suggest ppGpp accumulation and oxidative stress response - TPR protein and ABC transporter could be associated with *R. conorii* evasion	Renesto et al. (2008)
Rickettsia prowazekii	GC	Virulent strain compared to avirulent strain	- Genomic content is highly conserved - 1 gene is deleted and codes for a protein that could be implicated in apoptosis cascade - Virulent strain has paralogs for virulence factors	Ge et al. (2003) and Ge et al. (2004)
	GE	Response to heat shock stress	- Upregulation of genes coding for protein folding such as protease and chaperon	Audia et al. (2008)
Rickettsia rickettsii	GE	Virulent strain compared to avirulent strain	- Only four genes with unknown function are differentially expressed between avirulent and virulent strains	Ellison et al. (2008)
	GE	Response to iron limitation, shift of temperature, different types of cells and cold shock stress	- *Iron starvation, different cell types, small temperature shift*: Marginal differences - *Cold shock*: chaperone proteins (groELS and dnaK) downregulated as many other genes. This regulation is reverted after the arrest of the stress - *Prolongated Cold shock*: ABC transporter and alarmone pathway (spoT, ADP/ATP carrier protein) upregulated similar to starvation condition	Ellison et al. (2009)
Rickettsia typhi	GE	Response of shift of temperature	- Downregulation of genes that correlate with a slower growth rate in flea vector - Upregulation of genes that can maintain homeostasis	Dreher-Lesnick et al. (2008)
Tropheryma whipplei	GC	Comparison of 15 isolates	- Genomic content is highly conserved - WISP proteins are divergent among isolates - WNP repeat in WISP protein could be responsible of the *T. whipplei* genome plasticity	La et al. (2007a)
	GE	Response to heat and cold shock stresses	- *Heat shock*: upregulation of genes implicated in protein folding - *Cold shock*: upregulation of genes implicated in protein folding, adaptation of the membrane and accumulation of ppGpp	Crapoulet et al. (2006)
	GE	Response to doxycycline exposition	- No heat shock protein expressed - Inhibition of the translation which is doxycyline target and suggests growth arrest - Upregulation of ABC transporter that could extrude doxycycline	Van et al. (2007)

GC: genomic comparison.
GE: gene expression.

genes that code for proteins associated with post-translational modifications.

Rocky Mountain spotted fever (RMSF) is a tick-borne disease caused by *Rickettsia rickettsii* (Table 2). *Dermacentor andersoni*, the Rocky mountain wood tick, is the main known vector for RMSF (Walker, 1989; McDade and Newhouse, 1986). *R. rickettsii* is transmitted by the bite of an infected tick and may be maintained in nature, across several tick generations, through transovarial passage and transstadial passage. Ticks act as both the reservoir and vector of the pathogen (Dantas-Torres, 2007). *R. rickettsii* is a Gram-negative intracellular alpha-proteobacterium, belongs to the spotted fever group from the *Rickettsia* genus and multiplies freely in the cytoplasm or nucleus of host vascular endothelial cells. The 1.26 Mbp *R. rickettsii* genome (Sheila Smith strain) was sequenced and codes for 1345 ORFs that represent 76% of the coding content. Similar to the previously discussed intracellular bacteria, *R. rickettsii* has a small genome and an evolutionary tendency towards genomic reduction (Ellison et al., 2008). In a recent study, Ellison et al., 2009) performed a transcriptional analysis to determine the capacity of *R. rickettsii* to respond to environmental changes. The conditions, temperature shift, iron availability, cell type and cold shock, correspond to conditions encountered during the life of the bacteria in its arthropod vector and mammalian host. In contrast to slight shifts of temperature (34 °C to 22 °C or 37 °C to 25 °C), iron limitation or host cell species exposure, when *R. rickettsii* is exposed to an extreme shift in temperature, corresponding to cold shock, several genes are differentially expressed. Surprisingly, genes coding for chaperone proteins such as groELS and dnaK are downregulated upon cold shock, suggesting that the overexpression of these transcripts is associated with growth at physiologic temperatures. All genes that displayed regulation after a shift from 3 days of growth at 34 °C to 4 °C for 2 h were restored to their original levels of expression after being shifted back to 34 °C from 4 °C. Moreover, upon prolonged exposure to cold shock, the alarmone pathway appears to be upregulated by the overexpression of spoT, ADP/ATP carrier proteins and also a gene encoding an ABC transporter protein. These findings suggest that prolonged exposure to 4 °C may induce a response similar to exposure to starvation conditions. This study highlights that only minimal variations in the transcriptional profile appear during a small temperature shift, iron limitation or exposure to different cell types (Ellison et al., 2009), as also previously reported when comparing virulent and avirulent strains of *R. rickettsii* (Ellison et al., 2008). However, drastic environmental changes, such as cold shock, induce an adaptive transcriptional response from the bacteria that is similar to the response to starvation conditions.

Murine or endemic typhus is a flea-borne disease caused by *Rickettsia typhi* (Table 2). *Xenopsylla cheopis*, the rat flea, is the main known vector for murine typhus (Civen and Ngo, 2008). Murine typhus is a re-emerging febrile illness that is endemic in coastal areas throughout the world. Like other rickettiosis, it is characterized by a rash, fever and headaches. Murine typhus is acquired through flea a feces when *X. cheopsis* is feeding. *R. typhi* is a Gram-negative intracellular alpha-proteobacterium, belongs to the typhus group from the *Rickettsia* genus and multiplies freely in the endothelial cell cytoplasm of vessels (Raoult and Roux, 1997). The 1.11 Mbp *R. typhi* genome was sequenced in 2004 and codes for 877 ORFs that represent 76% of the coding content. As an intracellular bacterium, *R. typhi* has a small genome and an evolutionary tendency towards genomic reduction (McLeod et al., 2004). A transcriptional study was performed to examine how *R. typhi* genes are modulated upon a shift from 37 °C to 28 °C, mimicking the transition from mammalian host to the flea vector, using a short oligonucleotide *R. prowazekii* microarray (Dreher-Lesnick et al., 2008). The expression of genes involved in translation, transcription, post-translational modification and protein turnover and amino acid transfer and metabolism is downregulated following the temperature shift. The reduction in metabolic, transcriptional and translational activity correlates with a slower growth rate. However, the expression of a number of gene categories of particular interest, including those involved in intracellular trafficking and secretion, cell wall and membrane biogenesis and genes not fully characterized or with unknown function, was upregulated following the temperature shift. These genes may be important to maintain homeostasis in the flea vector. Several genes encoding proteins for the Type IV secretion system were also upregulated. Even though the role of Type IV secretion was not been established in *Rickettsiae*, it is known that Type IV secretion mediates the secretion of virulence factors essential for the intracellular survival of other bacteria. Based on this study, we can speculate that *R. typhi* modulates the transcription machinery to maintain homeostasis and survive in changing environments, such as being exposed to different temperatures that can mimic the arthropod vector to mammalian host transition.

The global transcriptional response of the *T. whipplei* response to temperature stress was analyzed using the same microarray (Crapoulet et al., 2006). The transcriptional profiles of *T. whipplei* exposed either to heat shock (43 °C) or to cold shock (28 °C and 4 °C) were compared to bacteria grown at 37 °C. The increase in temperature from 37 °C to 43 °C over 15 min resulted in the upregulation of genes encoding for several protein chaperones and one protease. Using bioinformatics, these genes have been proposed to be organized as an operon. Surprisingly, this thermal stress is not associated with the downregulation of ribosomal genes, suggesting that the increased temperature does not result in an arrest of *T. whipplei* growth. Upon cold shock, all genes that were regulated at 28 °C were highly regulated at 4 °C. The severe cold conditions resulted in the upregulation of genes encoding for protein chaperones. The regulation of genes related to fatty acid pathways was observed, highlighting an adaptation of *T. whipplei* to cold shock. The regulation of the alarmone pathway occurs during temperature stress, resulting in the accumulation of (p)ppGpp in the bacteria. Overall, this study demonstrates that *T. whipplei* modulates the expression of a number of genes to live under environmental stress. The molecular basis of *T. whipplei* antibiotic susceptibility was examined also by transcriptional analysis (Van et al., 2007). *T. whipplei* was treated with 0.5 and 5 μg/ml of doxycycline (corresponding to 1× MIC and 10× MIC), allowing for the detection of direct and indirect effects of the antibiotic, respectively. Surprisingly, in contrast to what was found in other microorganisms, the genes encoding heat shock proteins were not differentially expressed. As expected, however, several genes associated with translation were differentially expressed, which is consistent with the inhibitory effect of doxycycline on translation. The downregulation of genes associated with DNA replication suggests a growth arrest. Finally, genes encoding ABC transporters were upregulated following exposure to 10× MIC doxycycline. An ABC transporter pump could serve to clear the doxycycline from the bacterium to favor detoxification. The transcriptional mechanisms used to eliminate the antibiotic could explain why doxycycline has only bacteriostatic activity on *T. whipplei* (Boulos et al., 2005; Boulos et al., 2004).

Transcriptional analyses have been recently used to help construction of an axenic medium for *C. burnetii*. Previously, Omsland et al. (2008) have developed a nutrient medium called Complex Coxiella Medium. While growth was not observed in CCM, *C. burnetii* demonstrated metabolic activity in this medium, suggesting that modification of the CCM could support *C. burnetii* replication. In this regard, the authors compared whole genome transcriptional profiles of organisms replicating in Vero cells to cells incubated in CCM for 24 h in order to identify the nutritional deficiencies of CCM that could preclude *C. burnetii* cell division (Omsland et al., 2009). The strategy, which comprises microarray, genomic reconstruction and metabolic typing, used to establish axenic culture conditions for *C. burnetii*, might be applicable to improve media formulations that support the growth of other obligate intracellular bacteria.

4.2.3. In vitro analysis

Mediterranean spotted fever (MSF) is a tick-borne disease caused by *R. conorii* (Table 2). *Rhipicephalus sanguineus*, the brown dog tick, is the

main known vector for *R. conorii* (Table 1). MSF is mostly endemic in the Mediterranean area and is characterized by rash and fever. Infection of humans with *R. conorii* leads to vasculitis and lesions at the site of the tick bite. Eschars (tâche noire) result from injury to many small vessels caused by the MSF agent (Rovery and Raoult, 2008). *R. conorii* is a Gram-negative intracellular alpha-proteobacterium, belongs to the spotted fever group from the *Rickettsia* genus and multiplies in endothelial cells of small to medium vessels. After phagocytosis and internalization by the host cell, the phagocytic vacuole is rapidly lysed and *R. conorii* escape phagocytic digestion to multiply freely in the host cell cytoplasm and nucleus (Raoult and Roux, 1997). The approximately 1.26 Mbp *R. conorii* genome was sequenced in 2001 and codes for 1374 ORFs that represent 81% of the coding content. As an intracellular bacterium, *R. conorii* has a small genome and an evolutionary tendency for genomic reduction (Ogata et al., 2001).

In a recent study, we have explored the transcriptional profiles of *R. conorii* from a collection of eschars collected from MSF patients (Renesto et al., 2008) using a whole genome oligonucleotide microarray. *R. conorii* transcripts in eschars were compared with bacteria grown in a Vero cell monolayer. We observed that the transcriptional profile is highly conserved among the different eschars. We also highlighted the evidence for growth arrest within the eschars. The transcripts of genes involved in metabolism or associated with bacterial growth (translation, transcription, cell wall biogenesis, energy production, transport of carbohydrates, amino acids and nucleotides) were mainly down-regulated. Such decreases in global gene expression have been observed in a few bacteria during *in vitro* experiments and these changes in expressions could potentially thwart an immune response (Garzoni et al., 2007; Eymann et al., 2002). *R. conorii* downregulates also genes encoding rOmpA and rOmpB, that, are known to be two major antigenic determinants. In response to a possible immune attack from the host cell, *R. conorii* upregulates genes that are involved in genome repair. This may be related to the fact that DNA lesions can occur during oxidative stress, as seen in *Helicobacter pylori* (O'Rourke et al., 2003). The alarmone and purine pathways were upregulated, suggesting an accumulation of guanosine pentaphosphate or tetraphosphate ((p)ppGpp), which acts as a marker of the stress response. (p)ppGpp accumulation has been identified earlier, when *R. conorii* was placed in starvation stress condition (La et al., 2007b). *R. conorii* appears to have evolved a strategy of osmo-adaptation to inhibit the effects of hyperosmolarity by the upregulation of proline and betaine transporters that are osmo-protectants. We also suggested a putative role for ABC transporters and a tetratricopeptide repeat-containing protein in the evasion of *R. conorii*.

4.2.4. Transcriptional response

The genomes of intracellular bacteria are small. In this regard, a reduced number of genes, compared to free-living bacteria, may decrease the capacity of these bacteria to adapt to environmental changes. But, microarray-based transcriptional analysis reveals that, while obligate intracellular bacteria have a minimal set of genes, they are capable of adapting specifically to stresses induced by environmental changes or host attacks. However, most of these transcriptional studies are *in vitro* experiments and mimic conditions to which bacteria are exposed during infection and reflect host–pathogen interactions. Renesto et al. have performed the only *in vivo* transcriptional study from clinical specimens. Despite the constraints posed by the use of these precious samples, this study shows that it is possible to obtain the transcriptional profile and microarray-based transcriptional studies should not be limited to *in vitro* assays.

For the transcriptional analysis, the microarrays are printed with probes that recognize only the putative coding sequences, with most of the time a single probe per gene. The whole genome tiling arrays contain the entire genomic sequence, which is present on the microarray. It is not limited to coding sequences but can also monitor intergenic sequences. Among obligate intracellular bacteria, most of the micro-array-based transcriptional analysis are performed using non-tiling array and thus focus only on the coding sequences. However, the tiling array study on *M. leprae* shows that not only intergenic sequences but also the pseudogenes may be expressed. Their expression could potentially be associated with biological functions of transcription regulation as with ncRNA in eukaryotes. We have observed that genes implicated in the stress response are organized as operons that can be encompassed within regulons, allowing for the co-expression or repression of specific genes to improve bacterial adaptation.

Analyses of transcriptional profiles suggest that bacteria have specific response to environmental changes (Table 3). Some environmental changes induce no response from the bacteria as shown with *C. trachomatis* and *R. rickettsii*. To thwart the host response to bacterial infection, bacteria limit their growth. The growth arrest seems to be a specific response to environmental stress. Misfolding induced by temperature variation can also occur. The response of bacteria to heat and cold stresses has been examined in *T. whipplei* (Crapoulet et al., 2006), *R. prowazekii* (Audia et al., 2008) and *R. rickettsii* (Ellison et al., 2009). Obligate intracellular bacteria respond to the temperature shift by changing the expression of chaperone proteins and proteases that regulate the folding of proteins. ABC transporters are regulated with different types of bacterial stress, as shown with antibiotic-induced stress in *T. whipplei* (Van et al., 2007), with skin infection with *R. conorii* (Renesto et al., 2008) and cold shock (Ellison et al., 2009). Finally, secretion could also have a key role in bacterial pathogenicity. The overexpression of components of the secretion system may allow bacteria to deliver eukaryotic-like proteins, such as ankyrin or tetratricopeptide, into host cells that can either mimic or disrupt the function of eukaryotic proteins (Pan et al., 2008).

5. Conclusions and future perspectives

In contrast to studies examining the transcriptional profiles elaborated by host cells during infection with intracellular bacteria (Hossain et al., 2006), only a few studies using microarray technology have focused on transcriptional analyses of obligate intracellular or facultative intracellular bacteria (La et al., 2008). The number of techniques available to improve microarray experiments has increased rapidly for studies that examine the bacterial side of the host–pathogen interaction. These techniques have provided insights as to how obligate intracellular bacteria adapt to the host cell environment by targeted genomic variations and transcriptional regulation of specific genes. To date, most of the data generated have been performed with *in vitro* models, in which conditions can be controlled easily and large quantities of material can be harvested. Because the results are influenced by the model system used, the data must be interpreted with caution. In this respect, efforts must be made to increase the number of *in vivo* studies that more accurately reflect host–pathogen interactions.

Coordinated virulence gene expression is critical for bacteria during the course of infection. One of the most elegant and innovative approaches using microarray technology is represented by the recent work of Cossart and colleagues (Toledo-Arana et al., 2009). These investigators analyzed *Listeria monocytogenes* whole genome transcriptional profiles using a high-resolution tiling array. Transcriptional profiles were obtained from wild type and mutant bacteria *in vitro*, *ex vivo* and *in vivo*. Decades of bacterial genetics, cell biology and *in vivo* studies have ranked facultative intracellular *L. monocytogenes* among best models to study host–pathogen interactions and intracellular parasitism. The investigators have uncovered far more diverse types of regulated RNAs than expected. Recently, small RNA (sRNA) has emerged as regulators of virulence (Toledo-Arana et al., 2007). Several sRNAs that can target other sRNA or genes involved in metabolism and associated with virulence were found to be expressed and also regulated in *in vivo* experiments. They also identified differentially transcribed long-overlapping 5′ and 3′ untranslated

regions (UTR) that could be involved in RNA translation regulation through a mechanism known as a riboswitch. Although sensing and regulation have been traditionally ascribed to proteins, it is becoming increasingly clear that RNAs can perform these activities by means of a riboswitch (Winkler, 2005; Coppins et al., 2007) or other described mechanism (Brantl, 2007). Several previously described non-coding RNAs exhibit the same genetic expression pattern as the virulence-associated genes (Toledo-Arana et al., 2009). These data suggest a previously unknown regulatory mechanism in bacteria and could explain some of the difficulties of correlating transcriptional and proteomic data. These new networks of regulation seem to play a key role in bacterial pathogenicity and may be discovered eventually in intracellular bacteria as previously made for *M. leprae* (Akama et al., 2009) or *C. trachomatis* (Albrecht et al., 2009),

As microarray technology advances, investigators are learning that the entire genome may contain information crucial to the host–pathogen interaction. In this respect, comparative genomic approaches, associated with gene expression analysis, may provide a more complete assessment of the complex interactions that occur between the pathogen and its host. The characterization of non-coding sequences as well as coding sequences should be achieved using high-resolution technologies that do not limit analysis to specific regions of the genome, such as tiling arrays (Toledo-Arana et al., 2009) (Akama et al., 2009) or high-throughput sequencing (Passalacqua et al., 2009; Albrecht et al., 2009). The comprehensive coordination and characterization of virulence genes or non-coding RNA expression in the course of an infection will undoubtedly increase our knowledge of host–pathogen interactions observed with intracellular bacteria.

Acknowledgements

The authors would like to thank Dr Christopher Paddock (CDC, Atlanta, USA) for critically reading the manuscript and Dr Eric Ghigo (URMITE, Marseille, France) for providing immunofluorescence pictures of intracellular bacteria in course of infection.

References

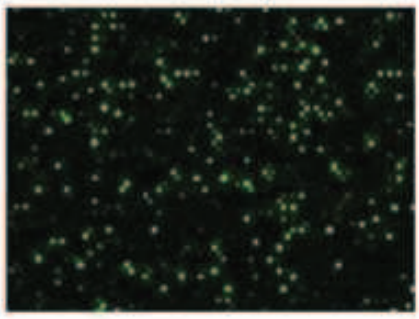
Akama, T., Suzuki, K., Tanigawa, K., Kawashima, A., Wu, H., Nakata, N., Osana, Y., Sakakibara, Y., Ishii, N., 2009. Whole-genome tiling array analysis of *Mycobacterium leprae* RNA reveals high expression of pseudogenes and noncoding regions. J. Bacteriol. 191, 3321.
Albrecht, M., Sharma, C.M., Reinhardt, R., Vogel, J., Rudel, T., 2009. Deep sequencing-based discovery of the *Chlamydia trachomatis* transcriptome. Nucleic Acids Res.
Alwine, J.C., Kemp, D.J., Stark, G.R., 1977. Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. Proc. Natl. Acad. Sci. U. S. A 74, 5350.
Anderson, B.E., Sims, K.G., Olson, J.G., Childs, J.E., Piesman, J.F., Happ, C.M., Maupin, G.O., Johnson, B.J., 1993. *Amblyomma americanum*: a potential vector of human ehrlichiosis. Am. J. Trop. Med. Hyg. 49, 239.
Andersson, J.O., Andersson, S.G., 1999. Insights into the evolutionary process of genome degradation. Curr. Opin. Genet. Dev. 9, 664.
Andersson, S.G., Zomorodipour, A., Andersson, J.O., Sicheritz-Ponten, T., Alsmark, U.C., Podowski, R.M., Naslund, A.K., Eriksson, A.S., Winkler, H.H., Kurland, C.G., 1998. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. Nature 396, 133.
Audia, J.P., Patton, M.C., Winkler, H.H., 2008. DNA microarray analysis of the heat shock transcriptome of the obligate intracytoplasmic pathogen *Rickettsia prowazekii*. Appl. Environ. Microbiol. 74, 7809.
Barker, L.P., 2006. *Mycobacterium leprae* interactions with the host cell: recent advances. Indian J. Med. Res. 123, 748.
Beare, P.A., Samuel, J.E., Howe, D., Virtaneva, K., Porcella, S.F., Heinzen, R.A., 2006. Genetic diversity of the Q fever agent, *Coxiella burnetii*, assessed by microarray-based whole-genome comparisons. J. Bacteriol. 188, 2309.
Beare, P.A., Howe, D., Cockrell, D.C., Omsland, A., Hansen, B., Heinzen, R.A., 2009. Characterization of a *Coxiella burnetii* ftsZ mutant generated by Himar1 transposon mutagenesis. J. Bacteriol. 191, 1369.
Bechah, Y., Capo, C., Mege, J.L., Raoult, D., 2008. Epidemic typhus. Lancet Infect. Dis. 8, 417.
Belland, R.J., Nelson, D.E., Virok, D., Crane, D.D., Hogan, D., Sturdevant, D., Beatty, W.L., Caldwell, H.D., 2003a. Transcriptome analysis of chlamydial growth during IFN-gamma-mediated persistence and reactivation. Proc. Natl. Acad. Sci. U. S. A. 100, 15971.
Belland, R.J., Zhong, G., Crane, D.D., Hogan, D., Sturdevant, D., Sharma, J., Beatty, W.L., Caldwell, H.D., 2003b. Genomic transcriptional profiling of the developmental cycle of *Chlamydia trachomatis*. Proc. Natl. Acad. Sci. U. S. A. 100, 8478.
Bentley, S.D., Maiwald, M., Murphy, L.D., Pallen, M.J., Yeats, C.A., Dover, L.G., Norbertczak, H.T., Besra, G.S., Quail, M.A., Harris, D.E., von, H.A., Goble, A., Rutter, S., Squares, R., Squares, S., Barrell, B.G., Parkhill, J., Relman, D.A., 2003. Sequencing and analysis of the genome of the Whipple's disease bacterium *Tropheryma whipplei*. Lancet 361, 637.
Binet, R., Maurelli, A.T., 2009. Transformation and isolation of allelic exchange mutants of *Chlamydia psittaci* using recombinant DNA introduced by electroporation. Proc. Natl. Acad. Sci. U. S. A. 106, 292.
Boulos, A., Rolain, J.M., Raoult, D., 2004. Antibiotic susceptibility of *Tropheryma whipplei* in MRC5 cells. Antimicrob. Agents Chemother. 48, 747.
Boulos, A., Rolain, J.M., Mallet, M.N., Raoult, D., 2005. Molecular evaluation of antibiotic susceptibility of *Tropheryma whipplei* in axenic medium. J. Antimicrob. Chemother. 55, 178.
Brantl, S., 2007. Regulatory mechanisms employed by cis-encoded antisense RNAs. Curr. Opin. Microbiol. 10, 102.
Brunelle, B.W., Nicholson, T.L., Stephens, R.S., 2004. Microarray-based genomic surveying of gene polymorphisms in *Chlamydia trachomatis*. Genome Biol. 5, R42.
Carlson, J.H., Hughes, S., Hogan, D., Cieplak, G., Sturdevant, D.E., McClarty, G., Caldwell, H.D., Belland, R.J., 2004. Polymorphisms in the *Chlamydia trachomatis* cytotoxin locus associated with ocular and genital isolates. Infect. Immun. 72, 7063.
Civen, R., Ngo, V., 2008. Murine typhus: an unrecognized suburban vectorborne disease. Clin. Infect. Dis. 46, 913.
Colston, M.J., 2001. *Mycobacterium leprae* genome sequence: a landmark achievement. Lepr. Rev. 72, 385.
Condon, C., 2007. Maturation and degradation of RNA in bacteria. Curr. Opin. Microbiol. 10, 271.
Coppins, R.L., Hall, K.B., Groisman, E.A., 2007. The intricate world of riboswitches. Curr. Opin. Microbiol. 10, 176.
Crapoulet, N., Barbry, P., Raoult, D., Renesto, P., 2006. Global transcriptome analysis of *Tropheryma whipplei* in response to temperature stresses. J. Bacteriol. 188, 5228.
Cummings, C.A., Relman, D.A., 2000. Using DNA microarrays to study host–microbe interactions. Emerg. Infect. Dis. 6, 513.
Dantas-Torres, F., 2007. Rocky Mountain spotted fever. Lancet Infect. Dis. 7, 724.
Di, C.F., Xie, Y., Paul-Satyaseela, M., Kim, K.S., 2005. Approaches to bacterial RNA isolation and purification for microarray analysis of *Escherichia coli* K1 interaction with human brain microvascular endothelial cells. J. Clin. Microbiol. 43, 4197.
Dreher-Lesnick, S.M., Ceraul, S.M., Rahman, M.S., Azad, A.F., 2008. Genome-wide screen for temperature-regulated genes of the obligate intracellular bacterium, *Rickettsia typhi*. BMC. Microbiol. 8, 61.
Dudoit, S., Gentleman, R.C., Quackenbush, J., 2003. Open source software for the analysis of microarray data. Biotechniques Suppl 45.
Ehrenreich, A., 2006. DNA microarray technology for the microbiologist: an overview. Appl. Microbiol. Biotechnol. 73, 255.
Eisen, M.B., Spellman, P.T., Brown, P.O., Botstein, D., 1998. Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. U. S. A. 95, 14863.
Ellison, D.W., Clark, T.R., Sturdevant, D.E., Virtaneva, K., Porcella, S.F., Hackstadt, T., 2008. Genomic comparison of virulent *Rickettsia rickettsii* Sheila Smith and avirulent *Rickettsia rickettsii* Iowa. Infect. Immun. 76, 542.
Ellison, D.W., Clark, T.R., Sturdevant, D.E., Virtaneva, K., Hackstadt, T., 2009. Limited transcriptional responses of *Rickettsia rickettsii* exposed to environmental stimuli. PLoS ONE 4, e5612.
Eriksson, S., Lucchini, S., Thompson, A., Rhen, M., Hinton, J.C., 2003. Unravelling the biology of macrophage infection by gene expression profiling of intracellular *Salmonella enterica*. Mol. Microbiol. 47, 103.
Eymann, C., Homuth, G., Scharf, C., Hecker, M., 2002. *Bacillus subtilis* functional genomics: global characterization of the stringent response by proteome and transcriptome analysis. J. Bacteriol. 184, 2500.
Fenollar, F., Raoult, D., 2003. Whipple's disease. Curr. Gastroenterol. Rep. 5, 379.
Fenollar, F., Puechal, X., Raoult, D., 2007. Whipple's disease. N. Engl. J. Med. 356, 55.
Fournier, P.E., Drancourt, M., Raoult, D., 2007. Bacterial genome sequencing and its use in infectious diseases. Lancet Infect. Dis. 7, 711.
Francois, P., Garzoni, C., Bento, M., Schrenzel, J., 2007. Comparison of amplification methods for transcriptomic analyses of low abundance prokaryotic RNA sources. J. Microbiol. Methods 68, 385.
Garzoni, C., Francois, P., Huyghe, A., Couzinet, S., Tapparel, C., Charbonnier, Y., Renzoni, A., Lucchini, S., Lew, D.P., Vaudaux, P., Kelley, W.L., Schrenzel, J., 2007. A global view of *Staphylococcus aureus* whole genome expression upon internalization in human epithelial cells. BMC. Genomics 8, 171.
Gaywee, J., Radulovic, S., Higgins, J.A., Azad, A.F., 2002. Transcriptional analysis of *Rickettsia prowazekii* invasion gene homolog (invA) during host cell infection. Infect. Immun. 70, 6346.
Ge, H., Chuang, Y.Y., Zhao, S., Ternenak, J.J., Ching, W.M., 2003. Genomic studies of *Rickettsia prowazekii* virulent and avirulent strains. Ann. N. Y. Acad. Sci. 990, 671.
Ge, H., Chuang, Y.Y., Zhao, S., Tong, M., Tsai, M.H., Ternenak, J.J., Richards, A.L., Ching, W.M., 2004. Comparative genomics of *Rickettsia prowazekii* Madrid E and Breinl strains. J. Bacteriol. 186, 556.
Goebl, M., Yanagida, M., 1991. The TPR snap helix: a novel protein repeat motif from mitosis to transcription. Trends Biochem. Sci. 16, 173.
Grifantini, R., Bartolini, E., Muzzi, A., Draghi, M., Frigimelica, E., Berger, J., Ratti, G., Petracca, R., Galli, G., Agnusdei, M., Giuliani, M.M., Santini, L., Brunelli, B., Tettelin, H., Rappuoli, R., Randazzo, F., Grandi, G., 2002. Previously unrecognized vaccine candidates against group B meningococcus identified by DNA microarrays. Nat. Biotechnol. 20, 914.
Hacia, J.G., 1999. Resequencing and mutational analysis using oligonucleotide microarrays. Nat. Genet. 21, 42.
Hackstadt, T., 1998. The diverse habitats of obligate intracellular parasites. Curr. Opin. Microbiol. 1, 82.

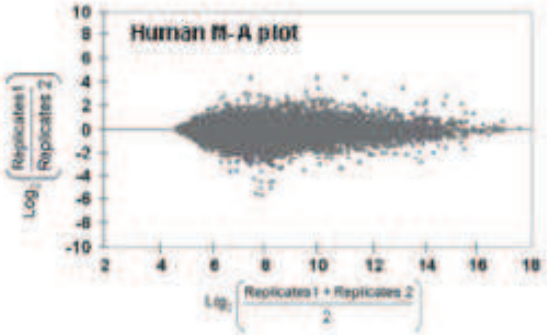
Hinton, J.C., Hautefort, I., Eriksson, S., Thompson, A., Rhen, M., 2004. Benefits and pitfalls of using microarrays to monitor bacterial gene expression during infection. Curr. Opin. Microbiol. 7, 277.
Hoover, T.A., Culp, D.W., Vodkin, M.H., Williams, J.C., Thompson, H.A., 2002. Chromosomal DNA deletions explain phenotypic characteristics of two antigenic variants, phase II and RSA 514 (crazy), of the *Coxiella burnetii* Nine Mile strain. Infect. Immun. 70, 6726.
Hossain, H., Tchatalbachev, S., Chakraborty, T., 2006. Host gene expression profiling in pathogen–host interactions. Curr. Opin. Immunol. 18, 422.
Hughes, T.R., Mao, M., Jones, A.R., Burchard, J., Marton, M.J., Shannon, K.W., Lefkowitz, S.M., Ziman, M., Schelter, J.M., Meyer, M.R., Kobayashi, S., Davis, C., Dai, H., He, Y.D., Stephaniants, S.B., Cavet, G., Walker, W.L., West, A., Coffey, E., Shoemaker, D.D., Stoughton, R., Blanchard, A.P., Friend, S.H., Linsley, P.S., 2001. Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. Nat. Biotechnol. 19, 342.
Jaluria, P., Konstantopoulos, K., Betenbaugh, M., Shiloach, J., 2007. A perspective on microarrays: current applications, pitfalls, and potential uses. Microb. Cell Fact. 6, 4.
Jenssen, T.K., Laegreid, A., Komorowski, J., Hovig, E., 2001. A literature network of human genes for high-throughput analysis of gene expression. Nat. Genet. 28, 21.
La, M.V., Crapoulet, N., Barbry, P., Raoult, D., Renesto, P., 2007a. Comparative genomic analysis of *Tropheryma whipplei* strains reveals that diversity among clinical isolates is mainly related to the WiSP proteins. BMC. Genomics 8, 349.
La, M.V., Francois, P., Rovery, C., Robineau, S., Barbry, P., Schrenzel, J., Raoult, D., Renesto, P., 2007b. Development of a method for recovering rickettsial RNA from infected cells to analyze gene expression profiling of obligate intracellular bacteria. J. Microbiol. Methods 71, 292.
La, M.V., Raoult, D., Renesto, P., 2008. Regulation of whole bacterial pathogen transcription within infected hosts. FEMS Microbiol. Rev. 32, 440.
Lathem, W.W., Crosby, S.D., Miller, V.L., Goldman, W.E., 2005. Progression of primary pneumonic plague: a mouse model of infection, pathology, and bacterial transcriptional activity. Proc. Natl. Acad. Sci. U. S. A. 102, 17786.
Leroy, Q., Rovery, C., Raoult, D., Renesto, P., 2009. Improvement of RNA purification from infected tissues to explore the *in vivo* host–pathogen interactions with microarrays. Clin. Microbiol. Infect.
Lin, J., Huang, S., Zhang, Q., 2002. Outer membrane proteins: key players for bacterial adaptation in host niches. Microbes. Infect. 4, 325.
Lipshutz, R.J., Fodor, S.P., Gingeras, T.R., Lockhart, D.J., 1999. High density synthetic oligonucleotide arrays. Nat. Genet. 21, 20.
Maurin, M., Raoult, D., 1999. Q fever. Clin. Microbiol. Rev. 12, 518.
McDade, J.E., Newhouse, V.F., 1986. Natural history of *Rickettsia rickettsii*. Annu. Rev. Microbiol. 40, 287.
McLeod, M.P., Qin, X., Karpathy, S.E., Gioia, J., Highlander, S.K., Fox, G.E., McNeill, T.Z., Jiang, H., Muzny, D., Jacob, L.S., Hawes, A.C., Sodergren, E., Gill, R., Hume, J., Morgan, M., Fan, G., Amin, A.G., Gibbs, R.A., Hong, C., Yu, X.J., Walker, D.H., Weinstock, G.M., 2004. Complete genome sequence of *Rickettsia typhi* and comparison with sequences of other rickettsiae. J. Bacteriol. 186, 5842.
Merhej, V., Royer-Carenzi, M., Pontarotti, P., Raoult, D., 2009. Massive comparative genomic analysis reveals convergent evolution of specialized bacteria. Biol. Direct. 4, 13.
Miura, K., Rikihisa, Y., 2007. Virulence potential of *Ehrlichia chaffeensis* strains of distinct genome sequences. Infect. Immun. 75, 3604.
Motley, S.T., Morrow, B.J., Liu, X., Dodge, I.L., Vitiello, A., Ward, C.K., Shaw, K.J., 2004. Simultaneous analysis of host and pathogen interactions during an *in vivo* infection reveals local induction of host acute phase response proteins, a novel bacterial stress response, and evidence of a host-imposed metal ion limited environment. Cell Microbiol. 6, 849.
Nicholson, T.L., Olinger, L., Chong, K., Schoolnik, G., Stephens, R.S., 2003. Global stage-specific gene regulation during the developmental cycle of *Chlamydia trachomatis*. J. Bacteriol. 185, 3179.
Nicholson, T.L., Chiu, K., Stephens, R.S., 2004. *Chlamydia trachomatis* lacks an adaptive response to changes in carbon source availability. Infect. Immun. 72, 4286.
Ochman, H., Moran, N.A., 2001. Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. Science 292, 1096.
Ogata, H., Audic, S., Renesto-Audiffren, P., Fournier, P.E., Barbe, V., Samson, D., Roux, V., Cossart, P., Weissenbach, J., Claverie, J.M., Raoult, D., 2001. Mechanisms of evolution in *Rickettsia conorii* and *R. prowazekii*. Science 293, 2093.
Omsland, A., Cockrell, D.C., Fischer, E.R., Heinzen, R.A., 2008. Sustained axenic metabolic activity by the obligate intracellular bacterium *Coxiella burnetii*. J. Bacteriol. 190, 3203.
Omsland, A., Cockrell, D.C., Howe, D., Fischer, E.R., Virtaneva, K., Sturdevant, D.E., Porcella, S.F., Heinzen, R.A., 2009. Host cell-free growth of the Q fever bacterium *Coxiella burnetii*. Proc. Natl. Acad. Sci. U. S. A. 106, 4430.
O'Rourke, E.J., Chevalier, C., Pinto, A.V., Thiberge, J.M., Ielpi, L., Labigne, A., Radicella, J.P., 2003. Pathogen DNA as target for host-generated oxidative stress: role for repair of bacterial DNA damage in *Helicobacter pylori* colonization. Proc. Natl. Acad. Sci. U. S. A. 100, 2789.
Paddock, C.D., Childs, J.E., 2003. *Ehrlichia chaffeensis*: a prototypical emerging pathogen. Clin. Microbiol. Rev. 16, 37.
Pan, X., Luhrmann, A., Satoh, A., Laskowski-Arce, M.A., Roy, C.R., 2008. Ankyrin repeat proteins comprise a diverse family of bacterial type IV effectors. Science 320, 1651.
Passalacqua, K.D., Varadarajan, A., Ondov, B.D., Okou, D.T., Zwick, M.E., Bergman, N.H., 2009. Structure and complexity of a bacterial transcriptome. J. Bacteriol. 191, 3203.
Picard, F., Robin, S., Lavielle, M., Vaisse, C., Daudin, J.J., 2005. A statistical approach for array CGH data analysis. BMC. Bioinformatics. 6, 27.
Pinkel, D., Segraves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W.L., Chen, C., Zhai, Y., Dairkee, S.H., Ljung, B.M., Gray, J.W., Albertson, D.G., 1998. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. Nat. Genet. 20, 207.
Raoult, D., Roux, V., 1997. Rickettsioses as paradigms of new or emerging infectious diseases. Clin. Microbiol. Rev. 10, 694.
Raoult, D., Lepidi, H., Harle, J.R., 2001. *Tropheryma whipplei* circulating in blood monocytes. N. Engl. J. Med. 345, 548.
Raoult, D., Ogata, H., Audic, S., Robert, C., Suhre, K., Drancourt, M., Claverie, J.M., 2003. *Tropheryma whipplei* Twist: a human pathogenic Actinobacteria with a reduced genome. Genome Res. 13, 1800.
Raoult, D., Marrie, T., Mege, J., 2005. Natural history and pathophysiology of Q fever. Lancet Infect. Dis. 5, 219.
Read, T.D., Brunham, R.C., Shen, C., Gill, S.R., Heidelberg, J.F., White, O., Hickey, E.K., Peterson, J., Utterback, T., Berry, K., Bass, S., Linher, K., Weidman, J., Khouri, H., Craven, B., Bowman, C., Dodson, R., Gwinn, M., Nelson, W., DeBoy, R., Kolonay, J., McClarty, G., Salzberg, S.L., Eisen, J., Fraser, C.M., 2000. Genome sequences of Chlamydia trachomatis MoPn and *Chlamydia pneumoniae* AR39. Nucleic Acids Res. 28, 1397.
Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., Volkert, T.L., Wilson, C.J., Bell, S.P., Young, R.A., 2000. Genome-wide location and function of DNA binding proteins. Science 290, 2306.
Renesto, P., Crapoulet, N., Ogata, H., La, S.B., Vestris, G., Claverie, J.M., Raoult, D., 2003. Genome-based design of a cell-free culture medium for *Tropheryma whipplei*. Lancet 362, 447.
Renesto, P., Rovery, C., Schrenzel, J., Leroy, Q., Huyghe, A., Li, W., Lepidi, H., Francois, P., Raoult, D., 2008. *Rickettsia conorii* transcriptional response within inoculation eschar. PLoS ONE 3, e3681.
Rosenow, C., Saxena, R.M., Durst, M., Gingeras, T.R., 2001. Prokaryotic RNA preparation methods useful for high density array analysis: comparison of two approaches. Nucleic Acids Res. 29, E112.
Rovery, C., Raoult, D., 2008. Mediterranean spotted fever. Infect. Dis. Clin. North Am. 22, 515 ix.
Rovery, C., La, M.V., Robineau, S., Matsumoto, K., Renesto, P., Raoult, D., 2005a. Preliminary transcriptional analysis of spoT gene family and of membrane proteins in *Rickettsia conorii* and *Rickettsia felis*. Ann. N. Y. Acad. Sci. 1063, 79.
Rovery, C., Renesto, P., Crapoulet, N., Matsumoto, K., Parola, P., Ogata, H., Raoult, D., 2005b. Transcriptional response of *Rickettsia conorii* exposed to temperature variation and stress starvation. Res. Microbiol. 156, 211.
Saeed, A.I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., Braisted, J., Klapa, M., Currier, T., Thiagarajan, M., Sturn, A., Snuffin, M., Rezantsev, A., Popov, D., Ryltsov, A., Kostukovich, E., Borisovsky, I., Liu, Z., Vinsavich, A., Trush, V., Quackenbush, J., 2003. TM4: a free, open-source system for microarray data management and analysis. Biotechniques 34, 374.
Saenz, H.L., Dehio, C., 2005. Signature-tagged mutagenesis: technical advances in a negative selection method for virulence gene identification. Curr. Opin. Microbiol. 8, 612.
Schachter, J., 1978. Chlamydial infections (third of three parts). N. Engl. J. Med. 298, 540.
Schena, M., Shalon, D., Davis, R.W., Brown, P.O., 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270, 467.
Schnappinger, D., Ehrt, S., Voskuil, M.I., Liu, Y., Mangan, J.A., Monahan, I.M., Dolganov, G., Efron, B., Butcher, P.D., Nathan, C., Schoolnik, G.K., 2003. Transcriptional adaptation of *Mycobacterium tuberculosis* within macrophages: insights into the phagosomal environment. J. Exp. Med. 198, 693.
Schneider, T., Moos, V., Loddenkemper, C., Marth, T., Fenollar, F., Raoult, D., 2008. Whipple's disease: new aspects of pathogenesis and treatment. Lancet Infect. Dis. 8, 179.
Selinger, D.W., Cheung, K.J., Mei, R., Johansson, E.M., Richmond, C.S., Blattner, F.R., Lockhart, D.J., Church, G.M., 2000. RNA expression analysis using a 30 base pair resolution *Escherichia coli* genome array. Nat. Biotechnol. 18, 1262.
Seshadri, R., Paulsen, I.T., Eisen, J.A., Read, T.D., Nelson, K.E., Nelson, W.C., Ward, N.L., Tettelin, H., Davidsen, T.M., Beanan, M.J., Deboy, R.T., Daugherty, S.C., Brinkac, L.M., Madupu, R., Dodson, R.J., Khouri, H.M., Lee, K.H., Carty, H.A., Scanlan, D., Heinzen, R.A., Thompson, H.A., Samuel, J.E., Fraser, C.M., Heidelberg, J.F., 2003. Complete genome sequence of the Q-fever pathogen *Coxiella burnetii*. Proc. Natl. Acad. Sci. U. S. A. 100, 5455.
Shalon, D., Smith, S.J., Brown, P.O., 1996. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. Genome Res. 6, 639.
Southern, E.M., 1992. Detection of specific sequences among DNA fragments separated by gel electrophoresis. 1975. Biotechnology 24, 122.
Stephens, R.S., Kalman, S., Lammel, C., Fan, J., Marathe, R., Aravind, L., Mitchell, W., Olinger, L., Tatusov, R.L., Zhao, Q., Koonin, E.V., Davis, R.W., 1998. Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. Science 282, 754.
Storz, J., Spears, P., 1977. Chlamydiales: properties, cycle of development and effect on eukaryotic host cells. Curr. Top. Microbiol. Immunol. 76, 167.
Talaat, A.M., Hunter, P., Johnston, S.A., 2000. Genome-directed primers for selective labeling of bacterial transcripts for DNA microarray analysis. Nat. Biotechnol. 18, 679.
Talaat, A.M., Ward, S.K., Wu, C.W., Rondon, E., Tavano, C., Bannantine, J.P., Lyons, R., Johnston, S.A., 2007. Mycobacterial bacilli are metabolically active during chronic tuberculosis in murine lungs: insights from genome-wide transcriptional profiling. J. Bacteriol. 189, 4265.
Tamas, I., Klasson, L.M., Sandstrom, J.P., Andersson, S.G., 2001. Mutualists and parasites: how to paint yourself into a (metabolic) corner. FEBS Lett. 498, 135.
Thomas, J.G., Olson, J.M., Tapscott, S.J., Zhao, L.P., 2001. An efficient and robust statistical modeling approach to discover differentially expressed genes using genomic expression profiles. Genome Res. 11, 1227.
Toledo-Arana, A., Repoila, F., Cossart, P., 2007. Small noncoding RNAs controlling pathogenesis. Curr. Opin. Microbiol. 10, 182.
Toledo-Arana, A., Dussurget, O., Nikitas, G., Sesto, N., Guet-Revillet, H., Balestrino, D., Loh, E., Gripenland, J., Tiensuu, T., Vaitkevicius, K., Barthelemy, M., Vergassola, M., Nahori, M.A.,

Soubigou, G., Regnault, B., Coppee, J.Y., Lecuit, M., Johansson, J., Cossart, P., 2009. The *Listeria* transcriptional landscape from saprophytism to virulence. Nature 459, 950.
Tuanyok, A., Tom, M., Dunbar, J., Woods, D.E., 2006. Genome-wide expression analysis of *Burkholderia pseudomallei* infection in a hamster model of acute melioidosis. Infect. Immun. 74, 5465.
Tusher, V.G., Tibshirani, R., Chu, G., 2001. Significance analysis of microarrays applied to the ionizing radiation response. Proc. Natl. Acad. Sci. U. S. A. 98, 5116.
Van Gelder, R.N., von Zastrow, M.E., Yool, A., Dement, W.C., Barchas, J.D., Eberwine, J.H., 1990. Amplified RNA synthesized from limited quantities of heterogeneous cDNA. Proc. Natl. Acad. Sci. U. S. A. 87, 1663.
Van, L.M., Barbry, P., Raoult, D., Renesto, P., 2007. Molecular basis of *Tropheryma whipplei* doxycycline susceptibility examined by transcriptional profiling. J. Antimicrob. Chemother. 59, 370.
Voth, D.E., Howe, D., Beare, P.A., Vogel, J.P., Unsworth, N., Samuel, J.E., Heinzen, R.A., 2009. The *Coxiella burnetii* ankyrin repeat domain-containing protein family is heterogeneous, with C-terminal truncations that influence Dot/Icm-mediated secretion. J. Bacteriol. 191, 4232.
Walker, D.H., 1989. Rocky Mountain spotted fever: a disease in need of microbiological concern. Clin. Microbiol. Rev. 2, 227.
Walker, S.L., Lockwood, D.N., 2007. Leprosy. Clin. Dermatol. 25, 165.
Walker, D.H., Ismail, N., Olano, J.P., McBride, J.W., Yu, X.J., Feng, H.M., 2004. *Ehrlichia chaffeensis*: a prevalent, life-threatening, emerging pathogen. Trans. Am. Clin. Climatol. Assoc. 115, 375.
Wang, E., Miller, L.D., Ohnmacht, G.A., Liu, E.T., Marincola, F.M., 2000. High-fidelity mRNA amplification for gene profiling. Nat. Biotechnol. 18, 457.
Weber, H., Polen, T., Heuveling, J., Wendisch, V.F., Hengge, R., 2005. Genome-wide analysis of the general stress response network in *Escherichia coli*: sigmaS-dependent genes, promoters, and sigma factor selectivity. J. Bacteriol. 187, 1591.
Winkler, H.H., 1987. Protein and RNA synthesis by isolated *Rickettsia prowazekii*. Infect. Immun. 55, 2032.
Winkler, W.C., 2005. Riboswitches and the role of noncoding RNAs in bacterial metabolic control. Curr. Opin. Chem. Biol. 9, 594.
Wood, D.O., Azad, A.F., 2000. Genetic manipulation of *Rickettsiae*: a preview. Infect. Immun. 68, 6091.
Yang, Y.H., Dudoit, S., Luu, P., Lin, D.M., Peng, V., Ngai, J., Speed, T.P., 2002. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Res. 30, e15.

Chapitre I

Improvement of RNA Purification from Infected Tissues to Explore the in vivo Host–Pathogen Interactions with Microarrays





Commentaire

La technologie des puces à ADN peut être considérée comme une méthode puissante pour augmenter les connaissances sur les interactions hôte-pathogène au cours de l'infection. Explorer simultanément les profils transcriptionnels du pathogène et de son hôte pourrait être une approche prometteuse et pourrait à terme favoriser le développement d'outils diagnostiques ou thérapeutiques [24].

Actuellement, la majorité des études utilisant les puces à ADN pour la caractérisation des interactions hôte-pathogène ciblent la réponse de l'hôte eucaryote du fait de limitations techniques pour les études transcriptionnelles des bactéries. De plus, ces études sont souvent basées sur des modèles *in vitro* et ne reflètent pas toujours les conditions *in vivo*. Les échantillons cliniques correspondant aux conditions *in vivo* étant souvent en quantité limitée, il serait donc encore plus difficile d'entreprendre à ce jour une étude simultanée des profils transcriptionnels de l'hôte et de son pathogène.

Pour pallier à cette problématique, nous avons mis au point une méthode permettant d'entreprendre simultanément l'analyse des profils transcriptionnels de l'hôte et de son pathogène à partir d'escarres provenant de patients souffrant de fièvre boutonneuse méditerranéenne. Cette méthode a eu pour but d'optimiser les étapes d'extraction et de séparation des ARN eucaryotes et procaryotes, afin d'obtenir ces derniers en qualité et en quantité suffisante pour réaliser par la suite des hybridations de puces à ADN.

Cet article montre que la stratégie développée permet de séparer les fractions d'ARN eucaryotes et procaryotes en quantité et qualité suffisante pour entreprendre des analyses transcriptionnelles par puce à ADN. De plus cette stratégie semble être adaptable à d'autres modèles d'interactions hôte-pathogène.

Article 1

Improvement of RNA Purification from Infected Tissues to Explore the *in vivo* Host–Pathogen Interactions with Microarrays

Quentin LEROY, Clarisse ROVERY, Didier RAOULT, Patricia RENESTO

Clinical Microbiology and Infection (CMI)

2009 Dec ;15 Suppl 2:114-5.

Unité des Rickettsies, IRD-CNRS UMR 6236, Faculté de Médecine,

Marseille, France

10.1111/j.1469-0691.2008.02189.x

Improvement of RNA purification from infected tissues to explore the in vivo host–pathogen interactions with microarrays

Q. Leroy, C. Rovery, D. Raoult and P. Renesto

Université de la Méditerranée, URMITE IRD-CNRS 6236, Faculté de Médecine, Bd Jean Moulin, Marseille, cedex, France

Rickettsiae are Gram-negative alpha proteobacteria and arthropod-borne disease agents of spotted fevers and typhus [1]. Because of their obligate intracellular lifestyle, the molecular mechanisms involved in their pathogenicity are still poorly understood. To circumvent the difficulties in working with these bacteria, rickettsial genomes have been sequenced, thus allowing further post-genomic studies.

DNA microarrays can be considered as powerful tools to understand the host–pathogen interactions in the course of infectious diseases. Exploring the RNA profiles of both host and pathogen promises to improve our knowledge of the infectious process, thus favouring the development of preventive or therapeutic strategies [2]. While published studies mainly focused on the eukaryotic response, the complementary picture, i.e. the bacterial response, was poorly analysed because of several technical limitations [3].

The transcriptomic analysis of *R. conorii* was recently performed by microarrays [4]. This bacterium is responsible for Mediterranean spotted fever (MSF), a disease characterised by a skin lesion called eschar caused by the bite of its vector, the brown dog tick, *Rhipicephalus sanguineus* [5]. The aim of this work was to analyse the transcriptome pattern of rickettsiae within such eschars using the same strategy based on the removal of eukaryotic contaminants coupled with subsequent random amplification of cDNA. Our goal was also to recover eukaryotic RNA from the same sample, thus considering both sides of the host–pathogen interaction.

Preliminary assays were performed using rabbit eschars generated by the sub-cutaneous inoculation of purified bacteria (not shown). Human eschars were collected by sterile scalpel excision from patients suffering from MSF and immediately placed at 4°C overnight in 1 mL of RNA Later (Qiagen, Valencia, California, USA) before storage at −80°C. The biopsies were rapidly decontaminated by a 5-min incubation in iodated alcohol and 1 min washing in nuclease free water. The tissues excised in small pieces were mechanically lysed with tungsten beads and using the Mixer Mill MM3 (Qiagen), before enzymatic digestion with proteinase K. The total RNA was extracted from resulting lysates using the RNeasy Micro kit (Qiagen). Eukaryotic RNA was removed from the total RNA sample using the MicroEnrich Kit (Ambion, Austin, Texas, USA) as described [4]. To recover the eukaryotic RNA from the same sample, the beads were then mixed in TE buffer pH 8 and heated for 1 min at 95°C. The supernatants were collected and precipitated with alcohol. The bacterial RNA was retro-transcribed using random primers and amplified with the Genomiphi kit (GE Healthcare, Piscataway, New Jersey, USA), while human RNA was amplified using poly d(T) primers with T7 polymerase (Agilent, Santa Clara, California, USA). To check for the quality of amplification, PCR were performed on cDNA using primers targeting the *rrs* gene from *R. conorii* (F- GGCTCAGAACGAACGCTATC/R-GTTAGCTGCGAAACCGAAAG) and the eukaryotic actin gene (F-GGACTTCGAGCAAGAGATGG/R-AGCACTGTGTTGGCGTACAG), respectively. Hybridisations of prokaryotic cDNA were achieved following Cy3 or Cy5 labelling of *R. conorii* from eschars and using bacteria grown in vero cells as control. *R. conorii* microarrays were obtained from Agilent ('Custom microarray' design) and processed as described [6]. Hybridisations of human microarrays (Agilent) were performed with one colour labelling as recommended.

The electrophoregrams obtained with the Bioanalyzer (Agilent) indicated that the extraction of

Corresponding author and reprint requests: P. Renesto, Université de la Méditerranée, URMITE IRD-CNRS 6236, Faculté de Médecine, 27 Bd Jean Moulin, 13385 Marseille, cedex 05, France
E-mail: patricia.renesto@univmed.fr

No conflicts of interest declared.

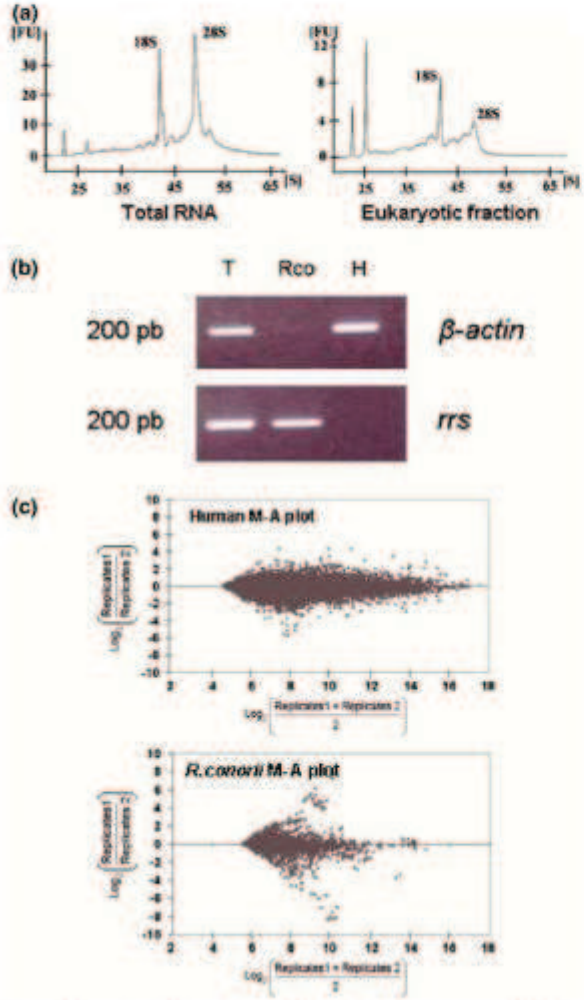


Fig. 1. Transcriptomic profiles of host-pathogens. (a) Electrophoregrams of RNA. (b) RT-PCR amplification using primers targeting β-actin or rrs. T (total RNA), Rco (purified *R. conorii*), H (human RNA recovered from beads). (c) M-A plot of biological replicates of *R. conorii* and human microarrays.

infected tissue preserved the integrity of the total RNA (Fig. 1a). Our protocol yielded around 500 ng of total RNA starting from 1 mg of infected eschars. Following the subtractive hybridisation approach, around 50 ng of purified prokaryotic RNA per mg of infected tissues were obtained. Up to 90% of the eukaryotic RNA retained onto beads was recovered while some degradation was likely to occur (Fig. 1a). The purification of both fractions was confirmed by RT-PCR. Thus, we successfully amplified β-actin (eukaryotic target) and *rrs* (prokaryotic target) from the initial sample of total RNA. In contrast, the purified prokaryotic fraction was found to be devoid of eukaryotic contaminants, while the RNA eluted from beads gave only a positive signal when using primers targeting actin gene (Fig. 1b). Using such a sample, reproducible patterns of microarray hybridisation were obtained. The M-A plots showed that the signals of labelled targets exhibited homogenous distribution (Fig. 1c).

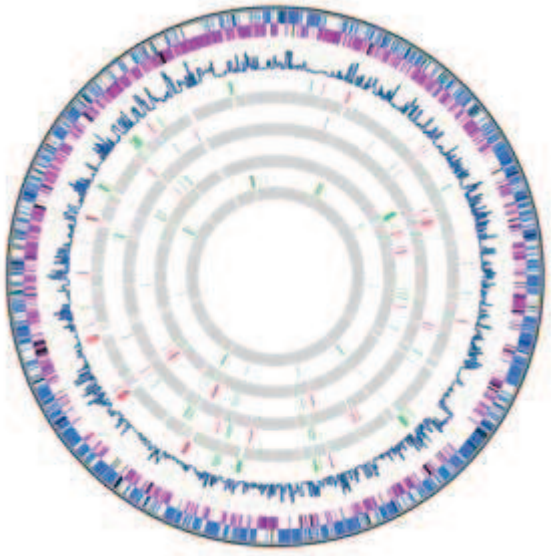
This approach has great potential to study the whole picture of host–pathogen interactions by microarrays.

REFERENCES

1. Raoult D, Roux V. Rickettsioses as paradigms of new or emerging infectious diseases. *Clin Microbiol Rev* 1997; **10:** 694–719.
2. Bryant PA, Venter D, Robins-Browne R, Curtis N. Chips with everything: DNA microarrays in infectious diseases. *Lancet Infect Dis* 2004; **4:** 100–111.
3. La MV, Raoult D, Renesto P. Regulation of whole bacterial pathogen transcription within infected hosts. *FEMS Microbiol Rev* 2008; **32:** 440–460.
4. La MV, Francois P, Rovery C *et al.* Development of a method for recovering rickettsial RNA from infected cells to analyze gene expression profiling of obligate intracellular bacteria. *J Microbiol Methods* 2007; **71:** 292–297.
5. Raoult D, Weiller PJ, Chagnon A, Chaudet H, Gallais H, Casanova P. Mediterranean spotted fever: clinical, laboratory and epidemiological features of 199 cases. *Am J Trop Med Hyg* 1986; **35:** 845–850.
6. Garzoni C, Francois P, Huyghe A *et al.* A global view of *Staphylococcus aureus* whole genome expression upon internalization in human epithelial cells. BMC. *Genomics* 2007; **8:** 171.

Chapitre II

Coxiella burnetii Transcriptional analysis Revealed Serendipity Clusters of Regulation in Intracellular Bacteria



Commentaire

Comme nous l'avons précédemment exposé, *Coxiella burnetii* est hautement résistante à différents stress environnementaux tels que les variations de pH, l'utilisation de produits désinfectants, la dessiccation mais aussi les variations de température [7]. Cette résistance permet à *C. burnetii* de survivre longuement dans les produits de parturition infectés lors d'avortements d'animaux d'élevage et permet la dissémination de bactéries viables [7] pouvant provoquer des larges épidémies [36]. Le changement de température est le stress environnemental auquel les organismes vivants sont le plus exposés [41]. Afin de surmonter les conditions critiques que peuvent générer de fortes variations de température, les bactéries ont développé des mécanismes complexes et spécifiques [41].

Les génomes des bactéries intracellulaires strictes étant généralement de petite taille, ils pourraient être associés à une faible capacité d'adaptation aux changements environnementaux. Cependant, différentes études portant sur des bactéries intracellulaires strictes à petit génome ont montré que ces dernières étaient capables d'avoir une réponse spécifique et adaptée lorsqu'elles subissaient une variation de température importante (type « cold » et « heat » shock). C'est le cas de *T. whipplei* [42] et de plusieurs espèces de rickettsies [43–45]. Ces réponses aux stress thermiques impliquaient la régulation de gènes codant pour des protéines chaperonnes mais aussi pour des protéines induisant l'arrêt de la croissance bactérienne, comme par exemple la synthèse de (p)ppGpp.

Pour répondre efficacement à ces conditions de stress, la bactérie doit être capable de coordonner l'expression de gènes spécifiques. Le dogme de l'expression génétique promoteur - facteur de transcription - opéron a été récemment mis à mal dans des approches globales de transcriptomique. En effet, plusieurs types d'ARN non codants (ARNnc) tels que les petits ARN (ARNs) ou les régions non traduites

des transcrits (UTR) ont ainsi été décrits comme pouvant jouer un rôle dans la régulation génétique, avec des phénomènes de « silencing » équivalant à ceux déjà caractérisés chez les eucaryotes [46]. De plus, ces régulations épigénétiques ont été associées à la virulence bactérienne. Cossart et *al.* ont identifié dans une récente étude utilisant la technologie des puces à ADN chez *Listeria monocytogenes,* des ARNnc régulés et associés à la virulence de la bactérie [47]. Une autre étude a aussi montré que l'ARN 6S de *Legionnella pneumophila* était impliqué dans la multiplication intracellulaire [48]. Enfin, dans une étude du RNome par RNAseq de *Chlamydia trachomatis* [49], il a été montré que la bactérie régulait des ARNnc comme des UTR ou des ARNs lors de son cycle cellulaire. Tous ces nouveaux niveaux de régulation transcriptionnelle ont mis en évidence que l'expression génétique était un monde plus complexe que le laisse penser le dogme originel.

Nous avons voulu durant cette thèse caractériser la réponse de *C. burnetii* au niveau transcriptionnel lors d'une exposition à de grandes variations de température (4°C et 43°C correspondant respectivement au « cold » et « heat » shock). Pour ce faire, nous avons construit une puce à ADN contenant plus de 95% des gènes de *C. burnetii,* en utilisant des stratégies initialement développées au sein du laboratoire pour l'analyse transcriptomique des bactéries intracellulaires.

Dans cet article, nous avons montré que la réponse transcriptionnelle précoce de *C. burnetii* à des changements de température est assez limitée. Cependant, nous avons remarqué que la réponse semble être équivalente entre une exposition à un « cold shock » ou un « heat shock ». De plus, les signatures transcriptionnelles observées pourraient correspondre à un passage à la forme pseudo-sporulée de *C. burnetii*. Ce modèle nous a permis de mettre en évidence un regroupement spatial des gènes différentiellement exprimés le long du génome que nous ensuite essayé de caractériser. Ce phénomène ne répondant pas aux dogmes promoteur - facteur de transcription – opéron et n'étant pas associé à un réseau biologique, nous posons l'hypothèse qu'il pourrait correspondre à un nouveau mode de régulation épigénétique qui reste encore à être caractérisé.

Article 2

Coxiella burnetii Transcriptional analysis Revealed Serendipity Clusters of Regulation in Intracellular Bacteria

Quentin LEROY[1], Kevin LEBRIGAND[2], Fabrice ARMOUGOM[1], Pascal BARBRY[2], Richard THIERY[3], Didier RAOULT[1]

Plos One

Accepté le 5 novembre 2010

1 - Unité des Rickettsies, IRD-CNRS UMR 6236, Faculté de Médecine, Marseille, France

2 – Institut de Pharmacologie Moléculaire et Cellulaire (IPMC), UMR 6079 CNRS-UNSA, Sophia Antipolis, France

3- Unité de pathologie des ruminants, Agence Française de Sécurité Sanitaire des Aliments (AFSSA), Sophia Antipolis France

Abstract

Coxiella burnetii, the causative agent of the zoonotic disease Q fever, is mainly transmitted to humans through an aerosol route. A spore-like form allows *C. burnetii* to resist different environmental conditions. Because of this, analysis of the survival strategies used by this bacterium to adapt to new environmental conditions is critical for our understanding of *C. burnetii* pathogenicity.

Here, we report the early transcriptional response of *C. burnetii* under temperature stresses. Our data show that *C. burnetii* exhibited minor changes in gene regulation under short exposure to heat or cold shock. While small differences were observed, *C. burnetii* seemed to respond similarly to cold and heat shock. The expression profiles obtained using microarrays produced in-house were confirmed by quantitative RT-PCR. Under temperature stresses, 190 genes were differentially expressed in at least one condition, with a fold change of up to 4. Globally, the differentially expressed genes in *C. burnetii* were associated with bacterial division, (p)ppGpp synthesis, wall and membrane biogenesis and, especially, lipopolysaccharide and peptidoglycan synthesis. These findings could be associated with growth arrest and witnessed transformation of the bacteria to a spore-like form. Unexpectedly, clusters of neighboring genes were differentially expressed. These clusters do not belong to operons or genetic networks; they have no evident associated functions and are not under the control of the same promoters. We also found undescribed but comparable clusters of regulation in previously reported transcriptomic analyses of intracellular bacteria, including *Rickettsia* sp. and *Listeria monocytogenes*.

The transcriptomic patterns of *C. burnetii* observed under temperature stresses permits the recognition of unpredicted clusters of regulation for which the trigger mechanism remains unidentified but which may be the result of a new mechanism of epigenetic regulation.

Introduction

C. burnetii is a Gram-negative intracellular γ-proteobacterium that causes Q fever, a zoonotic disease with a worldwide distribution [1]. Q fever can manifest as an acute or chronic illness. Acute Q fever is typically a self-limiting febrile illness during which pneumonia or hepatitis can occur, whereas chronic Q fever is a severe illness that may cause patients to present endocarditis, vascular infection, osteomyelitis and chronic hepatitis [1]. The major route of contamination with *C. burnetii* is as an aerosol. *C. burnetii* displays antigenic variation in its lipopolysaccharides (LPS) [2]. In phase I, the bacterium is highly infectious, and this corresponds to the natural phase found in animals, humans and arthropods, whereas phase II, which is not very infectious, presents truncated LPS and can be obtained after several passages in cell culture or embryonated eggs [1]. The *C. burnetii* genome was sequenced in 2003, and its size is approximately 2 Mbp with a plasmid of approximately 38 kbp [3]. Recently, five new isolates of *C. burnetii* were sequenced [4].

C. burnetii displays a complex intracellular cycle, leading to the formation of spore-like forms [5]. McCaul and Williams have proposed the terms “small-cell variant” (SCV) and “large-cell variant” (LCV) to differentiate the two *C. burnetii* cell forms observed in persistently infected cells [6]. SCVs are metabolically inactive and resistant to osmotic pressure and correspond to the extracellular form of the bacterium. SCVs attach to the eukaryotic cell membrane to enter phagocytic cells. After phagolysosomal fusion, acid activation of the metabolism of SCVs may lead to the formation of LCVs. Both LCVs and SCVs have a typical bacterial Gram-negative cell wall with two layers separated by the periplasmic space. However, a dense material fills the periplasmic space in SCVs. This material is composed of proteins and peptidoglycans and may explain the increased resistance of SCVs to environmental conditions [7]. The extracellular forms of *C. burnetii* resist environmental conditions such as desiccation and low or high pH, chemical products

such as ammonium chloride, disinfectants such as 0.5% sodium hypochlorite, and UV radiation [1,8].

Temperature change is the most common stress that all living organisms encounter in natural habitats. To overcome critical situations that could be generated by extreme temperatures, bacteria have evolved complex and specific mechanisms that are referred to as cold shock and heat shock responses [9]. Intracellular bacteria exhibit small genomes that show an evolutionary tendency toward genomic reduction, which could be associated with a lower adaptation capacity to environmental changes [10–12]. A number of intracellular bacteria have been observed to adapt to environmental changes, including *T. whipplei* and *Rickettsia* sp. [13–16]. Different obligate intracellular bacteria have exhibited the expression of specific genes encoding chaperone proteins and proteases that regulate the misfolding of proteins under stress conditions and alarmone accumulation. A previous transcriptional microarray study has been performed to improve an axenic medium for the *C.burnetii* culture [17].

Coordinated virulence gene expression is critical for bacteria during the course of infection. Global transcriptomic approaches have highlighted epigenetic mechanisms associated with bacterial pathogenicity. Cossart et al. showed that noncoding RNA (ncRNA) called small RNA (sRNA) was associated with *Listeria monocytogenes* pathogenicity through use of tilling microarray technology [18]. More recently, a sRNA microarray approach allowed researchers to discover that 6S RNA is implicated in intracellular multiplication [19]. A bacterial RNA seq study found that *Chlamydia trachomatis* exhibits regulation of ncRNA, including UTRs and sRNAs, during its cellular cycle [20]. These ncRNAs are involved in mechanisms that target gene regulation [21–24]. These levels of regulation show that bacterial gene regulation seems to be much more complicated than suggested by the promoter-and-transcription-factor paradigm.

Here, the early adaptive responses and the regulation mechanisms of *C. burnetii* exposed to various sudden temperature shifts were investigated using a whole-genome microarray. We also focus on the specific regulation mechanisms of *C. burnetii* and other intracellular bacteria to adapt in response to environmental stress.

Results

Microarray experiments

The complete transcriptional profile of *C. burnetii* exposed to stress temperatures was determined using a whole-genome microarray. Our microarray was spotted in quadruplicate and contained 1990 gene probes that corresponded to ~98.7% of the coding sequences of this species. Our microarray was validated by self-comparison with gDNA and cDNA hybridization (data not shown). In our experimental design, the reference group corresponded to the Nine Mile strain growing at 35°C in normal conditions, while the test group corresponded to the Nine Mile strain exposed to stress temperatures for 30 or 60 min. Bacteria were submitted to stress temperatures of 4 or 42°C, which represent the cold shock (CS) or heat shock (HS), respectively. RNA from bacteria and L929 cells were extracted simultaneously to avoid changes in transcriptomic profile after the bacterial purification process. Eukaryotic RNA was depleted using the MicrobEnrich Kit, which is based on a subtractive hybridization strategy. We found an atypical profile for *C. burnetii* RNA (Figure 1). The cDNA was amplified using random nucleotides and the highly processive phi29 polymerase. The hybridizations were performed in triplicate with three independent cultures. Quantification and *t*-test analyses were applied to determine the genes that were differently expressed at a significant level of confidence of above 95% with a 2-fold cut-off. To confirm the global response of the Nine Mile strain, RT-PCR was performed.

General overview

The differentially expressed genes and transcriptomic profile of *C. burnetii* grown at 35°C and then submitted either to heat shock (42°C) or to cold shock (4°C) for 30 min or 1 h, respectively, are shown in Table 1. Our transcriptomic analysis of the *C. burnetii* response to stress temperatures revealed the differential expression of 190 genes, including 140 genes for the CS treatment (85 for 30 min and 62 for 60 min of exposure) and 96 genes for the HS treatment (49 genes for 30 min and 58 for 60 min of exposure) (Supplementary data). The percentages of these genes

according to their functional classification based on COG are expressed in pie charts in Figure 2. Surprisingly, a clustering analysis of the differentially expressed genes under the four temperature stress conditions (Figure 2) showed that only small differences of expression were detectable between the four treatments (Figure 2).

Functional analysis

We functionally classified the differentially expressed genes according to COG [25]. We determined the proportion of different functional categories for each condition (Figure 3) and for genes that were differentially expressed in at least one of the stress conditions. The main category of differentially expressed genes was cell wall, membrane and envelope biogenesis (M), with up to 15% following 30 min heat shock. The genes associated with category M encoded outer membrane proteins or proteins involved in the synthesis of lipopolysaccharides, peptidoglycans and mureins. The second principal category observed were genes involved in amino acid transport and metabolism (F), which mostly included genes coding for transport system components (arginine and dipeptides). Nucleotide transport and metabolism (E) genes were also highly regulated, especially under cold shock treatment. In contrast to category E, the intracellular trafficking, secretion, and vesicular transport functional category (U) appeared to be heat shock–specific. Genes involved in cellular functions (transcription (K), replication (L) and translation (J)), representing principal and secondary metabolism, were also differentially expressed. Genes without an associated COG or with an unknown function–associated COG represented approximately 45% of the genes that were differentially expressed.

Regulon organization

Figure 4 shows that targeted large zones of the bacterial chromosome were simultaneously regulated under stress conditions. We analyzed the structure of these clusters of regulation. Each cluster contained between 5 and 11 genes, at least half of which were differentially expressed (Supplementary data). These clusters contained genes that are not necessarily organized into operons, and they can be found on both

genomic strands. To check whether this clustering pattern was statistically significant, we split the genome into windows of 5 to 11 genes and counted the number of differentially expressed genes in each. We included in the clusters of regulation all of the windows in which at least the half of genes were found differentially expressed. We found that these clusters of regulation contained differentially expressed genes that were significantly associated compared to a random distribution in the genome (Figure 5 and Supplementary data). Although the genes mostly occurred in complete operons, single ORFs and incomplete operons were also present in some clusters (Supplementary data). To elucidate the mechanism of these regulation clusters, we monitored gene functions within the clusters based on COG classification, but we did not find an enrichment of any specific functional category associated with our clusters compared to rest of the genome (Supplementary data). Finally, we focused on a functional protein association network using the STRING database 8.2 [26]. Based on the number of connections (score > 500) per protein, we determined whether the proteins encoded by the genes included in our clusters were specifically connected compared other *C. burnetii* proteins, but no significant differences were found (Supplementary data). Though these clusters of regulation included a number of genes that do not have obvious associated functions, we looked for networks that could link our clusters together and help us to understand this organization of gene expression regulation. Analysis of the protein association network showed that the different clusters seemed to be highly connected for the heat and cold shock conditions, but the connections were mostly spatial connections and not functional (Supplementary data). We also looked for structural genomic organization homology between *C. burnetii* and other sequenced γ-proteobacteria that are phylogenetically close to this species according to their 16S rRNA sequences (*Legionella* sp. and *Francisella* sp.). The genes implicated in clusters of regulation in *C. burnetii* presented no clearly identified synteny with those of *Legionella* sp. or *Francisella* sp. (Supplementary data). Finally, we compared the promoter sequences included in our clusters. We aligned the regions from -1000 bp to the translation start site (TSS). The phylogenic trees obtained from these alignments did not show

clustering of promoters associated with either up- or downregulation. We also examined predicted promoters using the Neural Network Promoter Prediction method [27]. We did not find any clearer clustering of promoters associated with gene regulation (Supplementary data). We also extracted the region from -10 to the TSS for every transcriptional unit and analyzed the CG% of these sequences (Supplementary data) to look for a correlation between GC% and transcriptional regulation. We observed no correlation between transcriptional regulation and the GC% of the -10 to translational start site sequences. Furthermore, we examined data from transcriptomic studies on other obligate intracellular bacteria. We collected data from the GEO database and Array Express to look for this kind of spatial regulation in other species, and we found that this type of regulation was also present in other species, including *Rickettsial* species *[14–16,28]*, *Tropheryma whipplei* [13] and *Listeria monocytogenes [18]* (Supplementary data). Figure 6 shows that large regions of the genomes of *R. rickettsii*, *T. whipplei* and *L. monocytogenes* can be highly regulated, comparable to the clusters of regulation found here.

Discussion

In this study, we examined the early response of gene expression patterns in *C. burnetii* to cold and heat shock using a global transcriptional approach based on microarray technology. Microarray-based transcriptional studies for obligate intracellular bacteria have limitations, such as obtaining RNA of sufficient quality and quantity [29,30]. Bacterial purification from infected cells involves several steps at 4°C, and bacteria are highly sensitive to cold shock [31]. To prevent the treatments from skewing the results, we extracted eukaryotic and prokaryotic RNA simultaneously, and eukaryotic RNA was then removed by subtractive hybridization. We observed an atypical rRNA profile (Figure 1) with three peaks. This atypical profile is due to an insertion sequence in the 23S rRNA gene, as previously described [32,33]. Considering the 23S rRNA split, we obtained purified *C. burnetii* RNA of good quality. This strategy of eukaryotic RNA depletion coupled with cDNA amplification was previously successfully devised for the global transcriptomic analysis of intracellular bacteria [34]. Finally, the results obtained by qRT-PCR validate our microarray hybridization experiments, which were carried out with bacterial RNA extracted from three independent experiments.

According to our results, *Coxiella burnetii* appears not to be highly sensitive to temperature shifts corresponding to CS and HS. We found few genes that were differentially expressed (around 3 ±1% per temperature stress conditions). The genes regulated upon exposure to stress temperatures showed minor changes, with up to a 4-fold change in their expression. Thus, we speculate that host cells provide a stable environment and can partially decrease transcriptional responses from occurring in obligate intracellular bacteria. Surprisingly, a clustering analysis of the differentially expressed genes under the four temperature stress conditions examined (Figure 2) shows that only slight differences of expression were detectable between the four conditions (as was previously shown with for *T. whipplei*) [13]. These similar transcriptomic profiles suggest that *C. burnetii* uses identical strategies to protect

itself from CS and HS during its early exposure to these conditions within cells (Figure 2).

Even if bacteria do have the capacity to adapt quickly, our study largely reflects the very early and early responses of *C. burnetii* to temperature shifts. Moreover, we point out these transcriptomic profiles reflect that *C. burnetii* could have been under growth arrest. A slowdown in cellular division in *C. burnetii* could be supported by the downregulation of genes coding for the septum placement (*ftsZ*) [35,36] [37], the segregation of the plasmid (*parB*) [38] and genes associated with cell division (*ftsY*, *gidAB*) [39,40]. The downregulation of genes implicated in alarmone degradation (*rpoZ*, *spoT and gmK*) [41,42] indicates a (p)ppGpp accumulation, which is involved in the stringent response and in bacterial sporulation [42–44]. The stringent response is classically followed by growth arrest. While most of genes coding for chaperone proteins are generally underexpressed, Hsp90 could be activated *via* HemE [45]. The cell wall and the membrane of *C. burnetii* seem to be modified and are associated with a spherical shape (*merBCD*) [43,46]. The bacteria also undergo homeostatic maintenance, in which ABC transport and efflux pumps are implicated (*artM*, *artQ*, *opp* system). The decrease in bacterial division coupled with the putative morphological aspects, the changes in the membrane and cell wall, and the homeostatic maintenance could potentially correlate to a transformation of *C. burnetii* into a metabolically inactive sporulation-like form (SCV) [5]. The SCV form seems to be associated with the stress response of *C. burnetii* and could confer on the bacteria strong resistance to environmental changes, such as CS and HS.

Surprisingly, we observed that significantly differentially expressed genes were mostly spatially clustered following exposure to stress temperatures (Figure 3), and we found that these genes were highly significantly spatially associated compared to a random distribution. Then, we hypothesized that this distribution was associated with a transcriptional regulation mechanism. Different levels of bacterial gene expression regulation have been previously characterized, such as organization

related to operons and regulons [47]. However, the clustering found in this study was not obviously associated with operon organization. Some genes were even found on the two different DNA strands. The second level of regulation could be related to functional associations and network connections. A study of the *Rickettsia prowazekii* transcriptional response to cold shock found that that only genes associated with posttranscriptional modification, such as protease and chaperon proteins, were differentially regulated [14]. However, our investigation of functional associations using COG classifications and network connections did not allow us to find any obvious associations. Furthermore, a study of genomic organization showed that our clusters of differentially expressed genes were not highly syntenic with those of other bacteria, in particular with phylogenetically closely related bacteria, including *L. pneumophila* and *F. tularensis*. This could indicate a lack of functional selection pressure. Another transcriptional level of regulation is the regulon. A regulon is a collection of genes or operons under regulation by the same regulatory protein. The observed downregulation of the gene coding for the RNA polymerase omega subunit in all of our experimental conditions directed our research toward the regulon phenomenon. We analyzed the downstream intergenic sequences of our differentially expressed genes to look for similarity in promoter patterns. As we mentioned previously, heat shock appears to be involved (p)ppGpp accumulation within these bacteria. ppGpp is known as a transcriptional regulator [48,49], and DksA, which binds to the RNA polymerase secondary channel, potentiates the effects of (p)ppGpp on transcription. The direct activation or repression of a gene promoter by (p)ppGpp and DksA is dictated by specific DNA sequence motifs [48,49]. Repressed genes are typically GC rich between the -10 hexamer box and the TSS, whereas activated genes are typically AT-rich in this position. Our analysis of promoter regions did not uncover any correlation of GC content and regulation within the regulated genes or the genes contained in clusters. The observed clustering of differentially expressed genes could not be attributed to (p)ppGpp or DksA regulation associated with the -10 to TTS region of these genes. The promoter analysis does not highlight a putative role associated with regulons. Thus, it is easy to speculate that the regulation

observed in this study could be due to epigenetic regulatory factors, or it could be an artifact from our methods.

To confirm the existence of this clustering of differentially expressed genes around the genome, we collected data from transcriptional microarray analyses of different obligate intracellular bacteria that we listed in a recent review [30]. From these data, we easily observed undescribed but comparable clusters of differentially expressed genes in different conditions for *Rickettsia* sp. [15,16,28,50] and *T. whipplei* [13] (Supplementary data and Figure 6). These studies have mostly focused on environmental changes. These findings indicate that regulation that can occur under conditions of stress. However, these studies were performed with low-density arrays and could be an artifact of the hybridization or analysis methods used. DNA probes are generally randomly spotted or synthesized on glass surfaces. In this regard, we can eliminate hybridization artifacts. A recent transcriptomic analysis of *L. monocytogenes* was performed using tilling microarrays [18]. Tilling arrays permit the investigation of whole genomes and should clearly reflect transcriptomic profiles. This transcriptomic analysis using tilling arrays for *L. monocytogenes* highlighted clusters of regulation (Supplementary data and Figure 6). It is possible that the limited number of RNA-seq studies of intracellular bacteria could explain why we have not observed these regulatory arrangements in previous RNA-seq studies. Here, we only focused on obligate intracellular bacteria and one facultative bacterium, but the observation of these regulatory clusters in different bacteria allowed us to confirm that these clusters appear to be a real, undescribed regulation phenomenon. Though we think that the definition of the observed regulatory clusters depends on the threshold applied to the data, this may be indicative of hot spots for coregulation, independent of operons or strain positions.

These hot spots of regulation do not correspond to classical transcriptomic regulation within the promoter-and-transcription-factor paradigm. We speculate that an epigenetic regulation mechanism is responsible for the clustering of differentially

expressed genes. Furthermore, recent studies based on bacterial RNA-seq methods [20,51] or tilling microarrays [18] have focused on a new level of gene expression regulation. ncRNA epigenetic regulation, including sRNA and riboswitches in 5' UTRs, have been highlighted and associated with bacterial virulence [18]. We can speculate that our differentially expressed genes could be targets of sRNA, and riboswitches could represent a plausible hypothesis to explain our observations. ABC transporters and efflux pumps are differentially expressed by *C. burnetii*. Riboswitches act as sensors and can activate or inhibit transcription in the presence of a specific molecule [52]. We can also hypothesize that there may be other epigenetic factors involved, such as hot spots of DNA methylation or DNA supercoiling, that could decrease the accessibility of transcription factors or RNA polymerase to promoter sequences [47]. Such phenomena are well known in eukaryotic models, such as in ncRNA silencing, and it is easy to speculate that this could be responsible for the clusters of regulation we have observed.

In conclusion, *C. burnetii* appears to be able to rapidly adapt itself to environmental changes such as cold and heat shock by altering the transcription of adapted genes that could be involved in transformation into a sporulation-like form. In bacteria, genes are organized into operons to facilitate the regulation of genes implicated in the same pathway. Here, we found that many of the genes that are differentially expressed upon exposure to temperature stresses are organized into clusters of regulation. Although we have not deciphered the mechanisms underlying these regulation clusters, this phenomenon seems to be widespread in obligate intracellular bacteria. Clustering related to the regulation of gene expression involved in bacterial adaptation could be advantageous for these bacteria. Thus, we will undertake new experiments related to transcriptional responses with longer exposure to stress conditions using technology that is adapted to highlight ncRNA or epigenetic factors (which we could not monitor with the microarray used here) to elucidate the phenomenon of gene regulation by clusters.

Materials and methods

Strain, medium and growth conditions

All experiments were performed with mid-log cultures of *C. burnetii* grown at 35°C on L929 cells in MEM (GIBCO, Invitrogen, Cergy-Pontoise, France) supplemented with 4% SVF (GIBCO) and 1% L-glutamine (GIBCO). For temperature stress experiments, flasks containing infected L929 cells were incubated at 4°C or 42°C for 30 min or 1 h with the Nine Mile I Strain. Infected cells were harvested using glass beads and centrifuged at 7,500 rpm for 10 min. Pellets were frozen using liquid nitrogen and stored at -80°C.

RNA extraction and purification

Pellets were resuspended in 100 µl of TE supplemented with 10 mg/ml of lysozyme (Euromedex, Souffelweyersheim, France) and incubated for 10 min at room temperature. Total RNA was extracted and purified from resuspended pellets using the RNeasy Mini Kit (Qiagen, Courtaboeuf, France) as recommended by the manufacturer. DNase treatment was performed using the DNA Turbo Free Kit (Ambion, Applied Biosystems, Courtaboeuf, France). Total RNA integrity was checked using the 2100 BioAnalyzer (Agilent Technologies, Palo Alto, CA), and the concentrations were quantified using the NanoDrop (Thermo, Wilmington, USA). Eukaryotic RNA and bacterial rRNA were depleted using the MicrobEnrich kit (Ambion) as previously described [53] and the MicrobExpress kit (Ambion), respectively. The integrity of bacterial RNA was checked using the 2100 BioAnalyzer, and the concentrations were quantified using the NanoDrop.

RNA labeling for microarray experiments

RNA was reverse-transcribed using M-MLV (Invitrogen, Cergy-Pontoise, France) and random hexamer primers (Invitrogen) as previously described. cDNAs were amplified using the processive polymerase phi29 with the GenomPhi illustrator V2 kit (GE HealthCare Lifescience, Orsay, France). This strategy was previously described [29]. The amplified cDNAs were labeled with the Bioprime Labeling

System (Invitrogen) using d-CTP Cy3/5 fluorochromes (GE HealthCare Lifescience). Labeled cDNAs were purified using QIAquick mini kit columns (Qiagen), and the level of incorporation was quantified using the NanoDrop.

Coxiella burnetii whole-genome microarray construction

OligoArray 2.0 [54,55] was used to design probes based on 2016 CDS extracted from the NC_002971.gb GenBank sequence file. OligoArray 2.0 integrates BLAST analysis against a nonredundant set of sequences and probe secondary structure analyses [56]. Oligonucleotide calculation parameters were set as follows: oligo length from 50- to 52-mers; GC percentage from 35 to 55%; melting temperature from 82 to 86°C. OligoArray 2.0 selected probes with the lowest cross-hybridization and an absence of secondary structure and balanced the set of probes in terms of melting temperature. Oligonucleotides containing five consecutive A, C, G or T were discarded. Following probes design, 1990 probes corresponding to 1990 distinct CDS where selected for synthesis. A total of 100 μmol of each probe were ordered from Sigma–Proligo (Paris, FRANCE) as 5′ amino–modified oligonucleotides. Oligonucleotide stocks were aliquoted for use in microarray fabrication. Oligonucleotides were diluted to a final concentration of 35–50 μM in 35% dimethyl sulfoxide, 100 mM potassium phosphate (pH 8.0). *C. burnetii* 2k microarrays (GEO reference GPL6675) were printed with a ChipWriterProarrayer (Bio-Rad, 1000 Alfred Nobel Drive Hercules, CA) on commercial HydroGel slides (Schott, Hattenbergstr 10 55122 Mainz, Germany) and processed according to the manufacturer's instructions.

Microarray hybridizations

Hybridization was carried out using two samples of labeled cDNA (75 pmol of each) that were labeled with Cy3 or Cy5 d-CTP. The pooled samples were hybridized using the GE Hybridization Kit (Agilent Technologies) as recommended by the manufacturer. The mixture was applied to a Surhyb 1 array (Agilent Technologies) and hybridized onto the *Coxiella burnetii* array using an Agilent hybridization

chamber (Agilent Technologies). Microarrays were hybridized for 17 h at 62°C in a rotating oven. Microarrays were washed using the GE washing buffers (Agilent Technologies) for 5 min with wash buffer 1 at room temperature followed by 1 min with wash buffer 2 at 37°C. Microarrays were dried using a bath of acetonitrile (VWR, Fontenay sous Bois, France). The microarrays were scanned using the microarray scanner C (Agilent Technologies) using XDR at 5 µm resolution.

Analysis of microarray data

All microarray results have been deposited in the GEO database (http://www.ncbi.nlm.nih.gov/geo/) under GEO series accession number GSE21778. The signal intensity and local background were measured for each spot by analyzing the array pictures with Feature Extractor software (Agilent Technologies). The data filtering and normalization were performed using Midas from the TM4 suite (TIGR). Data normalizations were performed using global normalization and Lowess normalization methods. Normalized data were processed using Tmev software from the TM4 suite (TIGR) with a *t*-test with a p-value of <0.05 and a cut-off of 2 for the fold change [50]. All experiments were conducted three times, which yielded 12 measurements per gene (representing four technical replicates in three biological replicates). The gene expression level was determined by determining the mean of the 12 values obtained for each probe.

Cluster of regulation analysis

Differentially expressed gene distributions were calculated using windows of 5 to 11 genes. The number of differentially expressed genes was counted in each window, and the distribution of differentially expressed genes was compared to a random distribution. For the synteny analyses, we compared *C. burnetii* to *Legionella* sp. and *Francisella* sp. using the Geneplot application. Geneplot is available on the NCBI website (http://www.ncbi.nlm.nih.gov). Functional analyses were performed using the Cluster of Orthologous Gene classification (COG) [25]. We used the operon organization algorithm available in MicrobesOnline to define transcriptional units

[57]. Protein network data were extracted from the program STRING version 8.2 [26]. We used interactions with a score > 0.5. We extracted all downstream intergenic sequences (-1000 to 0 bp) for all of the genes and considered as putative promoter sequences all intergenic sequences with a length > 50 bp. Promoter prediction has also been performed on putative promoter sequences using the Neural Network Promoter Prediction method [27]. Promoter predictions and downstream intergenic sequences corresponding to our differentially regulated genes were aligned using the Muscle 3.7 program [58]. Phylogenetic trees were built using MEGA 4 software [59]. We extracted all of the downstream intergenic sequences (-10 to 0 bp) for all of the transcriptional units and considered sequences with intergenic sequence lengths > 50 bp. We analyzed the CG% of the extracted sequences in comparison to the GC% of intergenic sequences. Statistical analyses were performed using GraphPad Prism version 5 software.

Real-Time RT-PCR

RNA was reverse-transcribed using M-MLV (Invitrogen) and random hexamer primers (Invitrogen) as recommended by the manufacturer. qPCR was performed on cDNAs for targeted transcripts using the Quantitec Probes Kit (Qiagen) with the 7900 HT PCR system (Applied Biosystems). The primers and probes used to perform qPCR were designed based on the five *C. burnetii* sequenced genomes available on the NCBI database. The sequences of primers and probes used are listed in the Supplementary data (Table S1). The relative expression ratios of target genes were determined by comparing housekeeping genes (*com1*, *16S*, *rpoB*) with differentially transcribed genes using the software of the 7900 HT qPCR system.

Acknowledgments

This work was supported by a grant from the Agence National pour la Recherche, France (contract number: ANR-06-PNRA-031; VIACOX project). We would like to thank V. Mehrej for bioinformatic support, V. Magnone for the microarray spotting, and P. Renesto and V. Duquesne for their input in this project.

Author Contributions

Conceived and designed the experiments: PB, RT, DR. performed the experiments: QL, KL, FA. Analyzed the data: QL, RT, DR. Contributed reagents/materials/analysis tools: PB, DR. Wrote the paper: QL, RT, DR.

Reference List

1. Maurin M, Raoult D (1999) Q fever. Clin Microbiol Rev 12: 518-553.

2. Hackstadt T, Peacock MG, Hitchcock PJ, Cole RL (1985) Lipopolysaccharide variation in Coxiella burnetti: intrastrain heterogeneity in structure and antigenicity. Infect Immun 48: 359-365.

3. Seshadri R, Paulsen IT, Eisen JA, Read TD, Nelson KE, Nelson WC, Ward NL, Tettelin H, Davidsen TM, Beanan MJ, Deboy RT, Daugherty SC, Brinkac LM, Madupu R, Dodson RJ, Khouri HM, Lee KH, Carty HA, Scanlan D, Heinzen RA, Thompson HA, Samuel JE, Fraser CM, Heidelberg JF (2003) Complete genome sequence of the Q-fever pathogen *Coxiella burnetii*. Proc Natl Acad Sci U S A 100: 5455-5460. 10.1073/pnas.0931379100 [doi];0931379100 [pii].

4. Beare PA, Unsworth N, Andoh M, Voth DE, Omsland A, Gilk SD, Williams KP, Sobral BW, Kupko JJ, III, Porcella SF, Samuel JE, Heinzen RA (2009) Comparative genomics reveal extensive transposon-mediated genomic plasticity and diversity among potential effector proteins within the genus Coxiella. Infect Immun 77: 642-656. IAI.01141-08 [pii];10.1128/IAI.01141-08 [doi].

5. McCaul TF, Williams JC (1981) Developmental cycle of Coxiella burnetii: structure and morphogenesis of vegetative and sporogenic differentiations. J Bacteriol 147: 1063-1076.

6. Wiebe ME, Burton PR, Shankel DM (1972) Isolation and characterization of two cell types of Coxiella burneti phase I. J Bacteriol 110: 368-377.

7. Amano K, Williams JC, McCaul TF, Peacock MG (1984) Biochemical and immunological properties of Coxiella burnetii cell wall and peptidoglycan-protein complex fractions. J Bacteriol 160: 982-988.

8. Heinzen RA, Hackstadt T, Samuel JE (1999) Developmental biology of Coxiella burnettii. Trends Microbiol 7: 149-154. S0966842X99014754 [pii].

9. Klinkert B, Narberhaus F (2009) Microbial thermosensors. Cell Mol Life Sci 66: 2661-2676. 10.1007/s00018-009-0041-3 [doi].

10. Andersson SG, Zomorodipour A, Andersson JO, Sicheritz-Ponten T, Alsmark UC, Podowski RM, Naslund AK, Eriksson AS, Winkler HH, Kurland CG (1998) The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. Nature 396: 133-140. 10.1038/24094 [doi].

11. Raoult D, Ogata H, Audic S, Robert C, Suhre K, Drancourt M, Claverie JM (2003) *Tropheryma whipplei* Twist: a human pathogenic Actinobacteria with a reduced genome. Genome Res 13: 1800-1809. 10.1101/gr.1474603 [doi];13/8/1800 [pii].

12. Ogata H, Audic S, Renesto-Audiffren P, Fournier PE, Barbe V, Samson D, Roux V, Cossart P, Weissenbach J, Claverie JM, Raoult D (2001) Mechanisms of evolution in *Rickettsia conorii* and *R. prowazekii*. Science 293: 2093-2098. 10.1126/science.1061471 [doi];293/5537/2093 [pii].

13. Crapoulet N, Barbry P, Raoult D, Renesto P (2006) Global transcriptome analysis of Tropheryma whipplei in response to temperature stresses. J Bacteriol 188: 5228-5239. 188/14/5228 [pii];10.1128/JB.00507-06 [doi].

14. Audia JP, Patton MC, Winkler HH (2008) DNA microarray analysis of the heat shock transcriptome of the obligate intracytoplasmic pathogen *Rickettsia prowazekii*. Appl Environ Microbiol 74: 7809-7812. AEM.00896-08 [pii];10.1128/AEM.00896-08 [doi].

15. Dreher-Lesnick SM, Ceraul SM, Rahman MS, Azad AF (2008) Genome-wide screen for temperature-regulated genes of the obligate intracellular bacterium, *Rickettsia typhi*. BMC Microbiol 8: 61. 1471-2180-8-61 [pii];10.1186/1471-2180-8-61 [doi].

16. Ellison DW, Clark TR, Sturdevant DE, Virtaneva K, Hackstadt T (2009) Limited transcriptional responses of *Rickettsia rickettsii* exposed to environmental stimuli. PLoS One 4: e5612. 10.1371/journal.pone.0005612 [doi].

17. Omsland A, Cockrell DC, Howe D, Fischer ER, Virtaneva K, Sturdevant DE, Porcella SF, Heinzen RA (2009) Host cell-free growth of the Q fever bacterium *Coxiella burnetii*. Proc Natl Acad Sci U S A 106: 4430-4434. 0812074106 [pii];10.1073/pnas.0812074106 [doi].

18. Toledo-Arana A, Dussurget O, Nikitas G, Sesto N, Guet-Revillet H, Balestrino D, Loh E, Gripenland J, Tiensuu T, Vaitkevicius K, Barthelemy M, Vergassola M, Nahori MA, Soubigou G, Regnault B, Coppee JY, Lecuit M, Johansson J, Cossart P (2009) The Listeria transcriptional landscape from saprophytism to virulence. Nature 459: 950-956. nature08080 [pii];10.1038/nature08080 [doi].

19. Faucher SP, Friedlander G, Livny J, Margalit H, Shuman HA (2010) Legionella pneumophila 6S RNA optimizes intracellular multiplication. Proc Natl Acad Sci U S A 107: 7533-7538. 0911764107 [pii];10.1073/pnas.0911764107 [doi].

20. Albrecht M, Sharma CM, Reinhardt R, Vogel J, Rudel T (2010) Deep sequencing-based discovery of the Chlamydia trachomatis transcriptome. Nucleic Acids Res 38: 868-877. gkp1032 [pii];10.1093/nar/gkp1032 [doi].

21. Argaman L, Hershberg R, Vogel J, Bejerano G, Wagner EG, Margalit H, Altuvia S (2001) Novel small RNA-encoding genes in the intergenic regions of Escherichia coli. Curr Biol 11: 941-950. S0960-9822(01)00270-6 [pii].

22. Bouvier M, Sharma CM, Mika F, Nierhaus KH, Vogel J (2008) Small RNA binding to 5' mRNA coding region inhibits translational initiation. Mol Cell 32: 827-837. S1097-2765(08)00807-1 [pii];10.1016/j.molcel.2008.10.027 [doi].

23. Frohlich KS, Vogel J (2009) Activation of gene expression by small RNA. Curr Opin Microbiol 12: 674-682. S1369-5274(09)00143-X [pii];10.1016/j.mib.2009.09.009 [doi].

24. Dambach MD, Winkler WC (2009) Expanding roles for metabolite-sensing regulatory RNAs. Curr Opin Microbiol 12: 161-169. S1369-5274(09)00011-3 [pii];10.1016/j.mib.2009.01.012 [doi].

25. Tatusov RL, Galperin MY, Natale DA, Koonin EV (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res 28: 33-36. gkd013 [pii].

26. Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, Doerks T, Julien P, Roth A, Simonovic M, Bork P, von MC (2009) STRING 8--a global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Res 37: D412-D416. gkn760 [pii];10.1093/nar/gkn760 [doi].

27. Reese MG (2001) Application of a time-delay neural network to promoter annotation in the Drosophila melanogaster genome. Comput Chem 26: 51-56. S0097-8485(01)00099-7 [pii].

28. Renesto P, Rovery C, Schrenzel J, Leroy Q, Huyghe A, Li W, Lepidi H, Francois P, Raoult D (2008) *Rickettsia conorii* transcriptional response within inoculation eschar. PLoS One 3: e3681. 10.1371/journal.pone.0003681 [doi].

29. La MV, Francois P, Rovery C, Robineau S, Barbry P, Schrenzel J, Raoult D, Renesto P (2007) Development of a method for recovering rickettsial RNA from infected cells to analyze gene expression profiling of obligate intracellular bacteria. J Microbiol Methods 71: 292-297. S0167-7012(07)00325-9 [pii];10.1016/j.mimet.2007.09.017 [doi].

30. Leroy Q, Raoult D (2010) Review of microarray studies for host-intracellular pathogen interactions. J Microbiol Methods 81: 81-95. S0167-7012(10)00067-9 [pii];10.1016/j.mimet.2010.02.008 [doi].

31. Weber MH, Marahiel MA (2003) Bacterial cold shock responses. Sci Prog 86: 9-75.

32. Afseth G, Mo YY, Mallavia LP (1995) Characterization of the 23S and 5S rRNA genes of Coxiella burnetii and identification of an intervening sequence within the 23S rRNA gene. J Bacteriol 177: 2946-2949.

33. Raghavan R, Miller SR, Hicks LD, Minnick MF (2007) The unusual 23S rRNA gene of Coxiella burnetii: two self-splicing group I introns flank a 34-base-pair exon, and one element lacks the canonical omegaG. J Bacteriol 189: 6572-6579. JB.00812-07 [pii];10.1128/JB.00812-07 [doi].

34. La MV, Crapoulet N, Barbry P, Raoult D, Renesto P (2007) Comparative genomic analysis of *Tropheryma whipplei* strains reveals that diversity among clinical isolates is mainly related to the WiSP proteins. BMC Genomics 8: 349. 1471-2164-8-349 [pii];10.1186/1471-2164-8-349 [doi].

35. Weiss DS (2004) Bacterial cell division and the septal ring. Mol Microbiol 54: 588-597. MMI4283 [pii];10.1111/j.1365-2958.2004.04283.x [doi].

36. Feucht A, Errington J (2005) ftsZ mutations affecting cell division frequency, placement and morphology in Bacillus subtilis. Microbiology 151: 2053-2064. 151/6/2053 [pii];10.1099/mic.0.27899-0 [doi].

37. Beare PA, Howe D, Cockrell DC, Omsland A, Hansen B, Heinzen RA (2009) Characterization of a Coxiella burnetii ftsZ mutant generated by Himar1 transposon mutagenesis. J Bacteriol 191: 1369-1381. JB.01580-08 [pii];10.1128/JB.01580-08 [doi].

38. Edgar R, Biek D, Yarmolinsky M (2006) P1 plasmid partition: in vivo evidence for the ParA- and ParB-mediated formation of an anchored parS complex in the absence of a partner parS. Mol Microbiol 59: 276-287. MMI4933 [pii];10.1111/j.1365-2958.2005.04933.x [doi].

39. Ogasawara N, Yoshikawa H (1992) Genes and their organization in the replication origin region of the bacterial chromosome. Mol Microbiol 6: 629-634.

40. Yim L, Moukadiri I, Bjork GR, Armengod ME (2006) Further insights into the tRNA modification process controlled by proteins MnmE and GidA of Escherichia coli. Nucleic Acids Res 34: 5892-5905. gkl752 [pii];10.1093/nar/gkl752 [doi].

41. Chatterji D, Ogawa Y, Shimada T, Ishihama A (2007) The role of the omega subunit of RNA polymerase in expression of the relA gene in Escherichia coli. FEMS Microbiol Lett 267: 51-55. FML532 [pii];10.1111/j.1574-6968.2006.00532.x [doi].

42. Chatterji D, Ojha AK (2001) Revisiting the stringent response, ppGpp and starvation signaling. Curr Opin Microbiol 4: 160-165. S136952740000182X [pii].

43. Bendezu FO, de Boer PA (2008) Conditional lethality, division defects, membrane involution, and endocytosis in mre and mrd shape mutants of Escherichia coli. J Bacteriol 190: 1792-1811. JB.01322-07 [pii];10.1128/JB.01322-07 [doi].

44. Traxler MF, Summers SM, Nguyen HT, Zacharia VM, Hightower GA, Smith JT, Conway T (2008) The global, ppGpp-mediated stringent response to amino acid starvation in Escherichia coli. Mol Microbiol 68: 1128-1148. MMI6229 [pii];10.1111/j.1365-2958.2008.06229.x [doi].

45. Saito M, Watanabe S, Yoshikawa H, Nakamoto H (2008) Interaction of the molecular chaperone HtpG with uroporphyrinogen decarboxylase in the cyanobacterium Synechococcus elongatus PCC 7942. Biosci Biotechnol Biochem 72: 1394-1397. JST.JSTAGE/bbb/80093 [pii].

46. Varma A, Young KD (2009) In Escherichia coli, MreB and FtsZ direct the synthesis of lateral cell wall via independent pathways that require PBP 2. J Bacteriol 191: 3526-3533. JB.01812-08 [pii];10.1128/JB.01812-08 [doi].

47. Martinez-Nunez MA, Perez-Rueda E, Gutierrez-Rios RM, Merino E (2010) New insights into the regulatory networks of paralogous genes in bacteria. Microbiology 156: 14-22. mic.0.033266-0 [pii];10.1099/mic.0.033266-0 [doi].

48. Potrykus K, Cashel M (2008) (p)ppGpp: still magical? Annu Rev Microbiol 62: 35-51. 10.1146/annurev.micro.62.081307.162903 [doi].

49. Dalebroux ZD, Svensson SL, Gaynor EC, Swanson MS (2010) ppGpp conjures bacterial virulence. Microbiol Mol Biol Rev 74: 171-199. 74/2/171 [pii];10.1128/MMBR.00046-09 [doi].

50. Bechah Y, El KK, Mediannikov O, Leroy Q, Pelletier N, Robert C, Medigue C, Mege JL, Raoult D (2010) Genomic, proteomic, and transcriptomic analysis of virulent and avirulent Rickettsia prowazekii reveals its adaptive mutation capabilities. Genome Res . gr.103564.109 [pii];10.1101/gr.103564.109 [doi].

51. Sharma CM, Hoffmann S, Darfeuille F, Reignier J, Findeiss S, Sittka A, Chabas S, Reiche K, Hackermuller J, Reinhardt R, Stadler PF, Vogel J (2010) The primary transcriptome of the major human pathogen Helicobacter pylori. Nature 464: 250-255. nature08756 [pii];10.1038/nature08756 [doi].

52. Henkin TM (2008) Riboswitch RNAs: using RNA to sense cellular metabolism. Genes Dev 22: 3383-3390. 22/24/3383 [pii];10.1101/gad.1747308 [doi].

53. La MV, Raoult D, Renesto P (2008) Regulation of whole bacterial pathogen transcription within infected hosts. FEMS Microbiol Rev 32: 440-460. FMR103 [pii];10.1111/j.1574-6976.2008.00103.x [doi].

54. Rouillard JM, Herbert CJ, Zuker M (2002) OligoArray: genome-scale oligonucleotide design for microarrays. Bioinformatics 18: 486-487.

55. Rouillard JM, Zuker M, Gulari E (2003) OligoArray 2.0: design of oligonucleotide probes for DNA microarrays using a thermodynamic approach. Nucleic Acids Res 31: 3057-3062.

56. Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31: 3406-3415.

57. Dehal PS, Joachimiak MP, Price MN, Bates JT, Baumohl JK, Chivian D, Friedland GD, Huang KH, Keller K, Novichkov PS, Dubchak IL, Alm EJ, Arkin AP (2010) MicrobesOnline: an integrated portal for comparative and functional genomics. Nucleic Acids Res 38: D396-D400. gkp919 [pii];10.1093/nar/gkp919 [doi].

58. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32: 1792-1797. 10.1093/nar/gkh340 [doi];32/5/1792 [pii].

59. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24: 1596-1599. msm092 [pii];10.1093/molbev/msm092 [doi].

Figure Legends

Figure 1: Eukaryotic RNA depletion and the atypical profile of *C. burnetii* rRNA
(A) This figure represents the electrophoregram showing the overlap of total RNA after RNA extraction and bacterial RNA after eukaryotic RNA depletion. (B) This figure represents the electrophoregram showing the overlap of bacterial RNA after eukaryotic RNA depletion RNA and bacterial mRNA after bacterial rRNA depletion. (C) This figure represents the gel-like representation of the fractions obtained after the different RNA depletions.

Figure 2: Hierarchical clustering of differentially expressed genes.
CS30, CS 60, HS30 and HS60 represent the cold shock stress for 30 min and 60 min and the heat shock stress for 30 min and 60 min, respectively. Green plots represent genes that are downregulated upon temperature stress, red plots represent genes that are upregulated upon temperature stress, and gray plots represent genes with variable regulation observed in biological replicates ($p>0,05$).

Figure 3: Functional category classification of genes differentially expressed upon exposure to stress temperatures
(A) The figure represents the proportion of genes differentially expressed according to the COG functional classification for each condition. (B) The pie chart represents the proportion of genes differentially expressed in at least one condition according to the COG.

Figure 4: Transcriptional profiles of the early responses to temperature stress.

Figure 5: Distribution of differentially expressed genes
This figure represents the total number of differentially expressed genes included in different window sizes (between 5 and 11 genes) that contain differentially expressed genes.

Figure 6: Transcriptional profiles from previously reported analyses of intracellular bacteria

Table 1: List of differentially expressed genes

Supplementary data

Figure S1: Venn diagram of differentially expressed genes

Figure S2: Clusters of regulation

Figure S3: *Coxiella burnetii* gene network connections

Figure S4: Differentially expressed gene networks

Figure S5: Promoter sequence analysis

Table S1: Table of primers and probes used for qRT-PCR and results

Table S2: Clusters of regulation within other obligate intracellular bacteria

Table S3: Statistical analysis of the distribution of differentially expressed genes

Table S4: Statistical analysis of functional categories within clusters

Table S5: Statistical analysis of network connections within clusters

Table S6: Synteny between *C. burnetii* and *Legionella* sp. and *Francisella* sp.

Table S7: Promoter -10 to start site analysis

Table S8: Gene information

NCBI accesion	Description	Gene name	COG	Fold Change			
				CS 30 min	CS 60 min	HS 30 min	HS 60 min
CBU_0029	1-acyl-sn-glycerol-3-phosphate acyltransferase	-	COG0204I	-2,08			
CBU_0034	putative acyl carrier protein	-	COG0236IQ			-2,21	-2,03
CBU_0044	hypothetical protein CBU_0044	-	-			-2,26	
CBU_0056	TolC family type I secretion outer membrane protein	-	COG1538MU . COG3266S			-2,90	
CBU_0080	not defined	-	not defined	2,01			
CBU_0081	prolyl-tRNA synthetase	*proS*	COG0442J	2,00		2,01	
CBU_0101	L-aspartate oxidase	*nadB*	COG0029H		2,06		
CBU_0111	2-amino-3-ketobutyrate coenzyme A ligase	*kbl*	COG0156H		2,08		
CBU_0134	not defined	-	not defined	-2,04			
CBU_0141	cell division protein	*ftsZ*	COG0206D			-2,28	
CBU_0168	acyl carrier protein	-	COG0236IQ		-2,01		
CBU_0169	Hypothetical protein (pseudo gene)	-	not defined		-2,20		
CBU_0170	Hypothetical protein (pseudo gene)	-	not defined		-2,08		
CBU_0188	Hypothetical membrane associated protein (pseudo gene)	-	not defined		-2,20		
CBU_0196	hypothetical protein CBU_0196	-	-	-2,83			
CBU_0275	uroporphyrinogen decarboxylase	*hemE*	COG0407H	-2,39			
CBU_0296	orotate phosphoribosyltransferase	*pyrE*	COG0461F	-2,43			

CBU_0297	exodeoxyribonuclease III	*xth*	COG0708L	-2,49			
CBU_0299	ribonuclease PH	*rph*	COG0689J	-2,75			
CBU_0300	hypothetical protein CBU_0300	-	COG1561S	-3,34			-2,01
CBU_0301	guanylate kinase	*gmk*	COG0194F	-2,83			
CBU_0302	DNA-directed RNA polymerase, omega subunit	*rpoZ*	COG1758K	-3,92	-2,03	-2,20	-2,26
CBU_0303	GTP pyrophosphokinase	*spoT*	COG0317TK	-2,25			
CBU_0304	putative endoribonuclease L-PSP	-	COG0251J	-2,62			-2,02
CBU_0360	not defined	-	not defined		2,03		2,38
CBU_0372	Fic family protein	-	COG3177S		2,07		
CBU_0373	MaoC domain-containing protein	-	COG2030I		2,02		
CBU_0379	anhydro-N-acetylmuramyl-tripeptide amidase	*ampD*	COG3023V		2,27		
CBU_0461	pyruvate dehydrogenase subunit E1	*aceE*	COG2609C	2,07		2,13	
CBU_0465	hypothetical protein CBU_0465	-	-			2,02	
CBU_0482	arginine ABC transporter periplasmic arginine-binding protein	-	COG0834ET		-2,03		-2,07
CBU_0483	arginine ABC transporter permease	*artQ*	COG4215E	-2,04	-2,29	-2,52	-2,39
CBU_0484	arginine transport system permease protein	*artM*	COG4160E		-2,32		
CBU_0488	Ser/Thr protein phosphatase family protein	-	COG0639T	-2,25	-2,18		
CBU_0489	phospholipase A1	-	COG2829M		-2,03		
CBU_0490	DNA polymerase X family/PHP domain-containing protein	-	COG1387ER, COG1796L	-2,35		-2,61	

CBU_0492	**putative glycerol-3-phosphate acyltransferase PlsX**	***plsX***	**COG0416I**	**-2,13**			
CBU_0598	**ADP-ribose diphosphatase NudE**	***nudE***	**COG0494LR**		**2,22**		
CBU_0599	**3'(2'),5'-bisphosphate nucleotidase**	***cysQ-1***	**COG1218P**		**3,12**		
CBU_0601	**Hypothetical protein (Pseudo gene)**	**-**	**not defined**		**2,30**		
CBU_0602	**Hypothetical protein (Pseudo gene)**	**-**	**not defined**		**2,16**		
CBU_0610	**hydroxymethylglutaryl-CoA reductase (NADPH)**	**-**	**COG1257I**	**-2,38**			
CBU_0611	**outer membrane protein assembly complex, YaeT protein**	***yaeT***	**COG4775M**	**-2,60**			
CBU_0612	**putative outer membrane chaperone protein Skp**	***ompH***	**COG2825M**	**-2,41**			
CBU_0613	**UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase**	***lpxD***	**COG1044M**	**-2,07**			
CBU_0614	**beta-hydroxyacyl-(acyl-carrier-protein) dehydratase FabZ**	***fabZ***	**COG0764I**	**-2,32**			
CBU_0615	**UDP-N-acetylglucosamine acyltransferase**	***lpxA***	**COG1043M**	**-2,28**			
CBU_0616	**hydrolase family protein**	**-**	**COG2179R**	**-2,33**			
CBU_0638	**branched-chain alpha-keto acid dehydrogenase subunit E2**	**-**	**COG0508C**		**-2,04**		
CBU_0642	**putative glycerophosphoryl diester phosphodiesterase**	**-**	**COG0584C**	**-2,04**	**-2,68**	**-2,00**	**-2,10**
CBU_0644	**plasmid stabilization system toxin protein**	**-**	**-**		**-2,07**		
CBU_0645	**prevent-host-death family protein**	**-**	**COG2161D**		**-2,24**		
CBU_0647	**3,4-dihydroxy-2-butanone 4-phosphate synthase domain...**	***ribA***	**COG0108H, COG0807H**	**-2,06**			

CBU_0648	6,7-dimethyl-8-ribityllumazine synthase	*ribH*	COG0054H	-2,20			
CBU_0654	hypothetical protein (pseudo gene)	-	not defined	-2,04		-2,04	
CBU_0671	mannose-1-phosphate guanylyltransferase	*rfbA*	COG0662G, COG0836M		2,21		
CBU_0674	SIS domain protein	-	COG0279G		2,48		
CBU_0700	sulfate adenylyltransferase/adenylylsulfate kinase	-	COG0529P, COG2046P	2,41		2,39	
CBU_0701	3'(2'),5'-bisphosphate nucleotidase	*cysQ-2*	COG1218P	2,12			
CBU_0702	short chain dehydrogenase/reductase family oxidoreductase	-	COG1028IQR	2,38		2,00	
CBU_0703	putative lipopolysaccharide/O-antigen ABC transporter...	-	COG1682GM	2,33		2,05	
CBU_0704	polysaccharide export ATP-binding protein	*rfbI*	COG1134GM			2,03	
CBU_0706	hypothetical protein CBU_0706	-	-	2,29		2,03	
CBU_0727	ABC transporter permease	-	COG0767Q		-2,05		-2,12
CBU_0729	ABC transporter periplasmic substrate-binding protein	-	COG1463Q				-2,14
CBU_0730	hypothetical protein CBU_0730	-	COG3218R				-2,37
CBU_0731	hypothetical protein CBU_0731	-	-				-2,37
CBU_0732	hypothetical protein (pseudo gene)	-	not defined				-2,57
CBU_0733	hypothetical protein (pseudo gene)	-	not defined				-2,08
CBU_0735	uridine phosphorylase	-	COG2820F		-2,15	-2,05	-2,37

CBU_0737	**trigger factor**	***tig***	**COG0544O**		**-2,04**		
CBU_0740	**ATP-dependent endopeptidase**	***lon***	**COG0466O**		**-2,09**		
CBU_0813	**hypothetical protein CBU_0813**	**-**	**-**		**-2,18**		
CBU_0816	**not defined**	**-**	**not defined**		**-2,03**		
CBU_0833	**ABC transporter permease/ATP-binding protein**	**-**	**COG1132V**	**2,31**		**2,04**	
CBU_0834	**putative methyltransferase**	**-**	**COG0500QR, COG2226H**	**2,72**		**2,26**	
CBU_0835	**hypothetical protein CBU_0835**	**-**	**COG2256L**	**2,66**	**2,07**	**2,27**	**2,32**
CBU_0836	**radical SAM domain-containing protein**	**-**	**COG0535R**	**2,31**		**2,09**	
CBU_0837	**hypothetical membrane spanning protein**	**-**	**-**	**2,12**			
CBU_0838	**lipopolysaccharide N-acetylglucosaminyltransferase**	**-**	**COG0438M**	**2,64**		**2,17**	**2,05**
CBU_0845	**UDP-glucose/GDP-mannose dehydrogenase family protein**	**-**	**COG0677M**				**2,15**
CBU_0846	**UDP-glucose 6-dehydrogenase**	***ugd***	**COG1004M**	**2,31**	**2,26**	**2,40**	**2,83**
CBU_0847	**oxidoreductase, short chain dehydrogenase/reductase family**	**-**	**COG1028IQR**				**2,11**
CBU_0856	**lipid ABC transporter permease/ATP-binding protein**	**-**	**COG1132V**				**2,11**
CBU_0857	**tetraacyldisaccharide 4'-kinase**	***lpxK***	**COG1663M**			**2,04**	
CBU_0861	**hypothetical protein CBU_0861**	**-**	**not defined**				**2,07**
CBU_0925	**putative lytic murein transglycosylase**	**-**	**COG2951M**		**2,03**		

CBU_0939	**AsmA family protein**	**-**	**COG2982M**	**2,07**			
CBU_0943	**rhodanese-like domain-containing protein**	**-**	**COG0607P**	**2,16**			
CBU_0945	**hypothetical membrane associated protein**	**-**	**-**	**2,70**			
CBU_0946	**death-on-curing family protein**	***rhuM***	**COG3654R, COG3943R**	**2,58**			
CBU_0948	**hypothetical protein CBU_0948**	**-**	**-**	**2,18**			
CBU_0949	**hypothetical cytosolic protein**	**-**	**-**	**2,38**			
CBU_0954	**phosphohydrolase (MutT/nudix family protein)**	**-**	**COG2816L**	**2,02**			
CBU_0959	**Bcr/CflA subfamily drug resistance transporter**	**-**	**COG0477GEPR**	**2,77**			
CBU_0964	**hypothetical protein CBU_0964**	**-**	**-**	**2,04**			
CBU_0991	**hypothetical cytosolic protein (pseudo gene)**	**-**	**not defined**				**-2,00**
CBU_1000	**lipoprotein releasing system, ATP-binding protein**	***lolD***	**COG1136V**		**-2,07**		
CBU_1032	**not defined**	**-**	**not defined**			**-2,16**	
CBU_1058	**hypothetical membrane spanning protein**	**-**	**COG3930S**	**2,22**			
CBU_1059	**RNA pseudouridine synthase family protein**	**-**	**COG1187J**	**2,15**			
CBU_1106	**not defined**	**-**	**not defined**	**2,05**			
CBU_1124	**chaperone-modulator protein CbpM**	**-**	**-**		**-2,08**		
CBU_1136	**enhanced entry protein enhC, tetratricopeptide repeat family**	**-**	**COG0790R**		**-2,05**		

CBU_1148	transcription-repair coupling factor	*mfd*	COG1197LK	2,06			
CBU_1150	not defined	-	not defined	2,25			
CBU_1151	HAD superfamily hydrolase	-	COG0546R	2,04			
CBU_1158	sterol delta-7-reductase	-	-	2,05			
CBU_1171	hypothetical protein CBU_1171	-	COG0432S				2,09
CBU_1202	hypothetical membrane spanning protein	-	-	-2,28		-2,10	-2,33
CBU_1203	glutamate synthase domain protein	-	COG0069E	-2,23		-2,10	-2,35
CBU_1204	succinate-semialdehyde dehydrogenase	-	COG1012C	-2,06			
CBU_1205	not defined	-	not defined	-2,16			-2,26
CBU_1206	delta(24(24(1)))-sterol reductase	-	COG2020O	-2,05			-2,09
CBU_1211	Hypothetical cytosolic protein (pseudo gene)	-	not defined	2,55		2,28	
CBU_1216	hypothetical protein (pseudo gene)	-	not defined	-2,35		-2,19	-2,51
CBU_1243	exodeoxyribonuclease VII large subunit	*xseA*	COG1570L				2,02
CBU_1263	hypothetical membrane associated protein	-	-				2,13
CBU_1266	lipoyl synthase	*lipA*	COG0320H				2,13
CBU_1267	DedA/PAP2 domain-containing protein	-	COG0586S, COG0671I		2,69		2,25
CBU_1268	pyridine nucleotide-disulphide oxidoreductase	-	COG0446R, COG1249C		2,10		

CBU_1269	thioesterase family protein	-	COG1607I		2,38		2,21
CBU_1272	hypothetical protein CBU_1272	-	-		2,07		2,23
CBU_1273	6-phosphofructokinase	-	COG0205G			2,05	2,09
CBU_1274	hypothetical protein CBU_1274	-	COG5513S				2,00
CBU_1282	carbamoyl-phosphate synthase small chain	*carA*	COG0505EF	-2,02			
CBU_1285	multidrug resistance protein B	-	COG0477GEPR	-2,14			
CBU_1286	hypothetical ATPase	-	COG1373R	-2,06			
CBU_1289	chaperone protein	*dnaJ*	COG0484O	-3,11		-3,01	
CBU_1290	chaperone protein	*dnaK*	COG0443O	-2,23			
CBU_1294	PilT domain-containing protein	-	COG1487R	-2,11			
CBU_1312	hypothetical protein CBU_1312	-	-			-2,13	
CBU_1326	threonyl-tRNA synthetase	*thrS*	COG0441J	2,00			
CBU_1331	CBS domain-containing protein	-	COG0517R	2,18			
CBU_1338	D-alanine--D-alanine ligase	*ddl*	COG1181M		-2,10		
CBU_1384	uridylate kinase	*pyrH*	COG0528F		2,24		
CBU_1385	elongation factor Ts	*tsf*	COG0264J		2,04		
CBU_1420	hypothetical protein (pseudo gene)	-	not defined				-2,09
CBU_1462	Hypothetical membrane associated protein (pseudo gene)	-	not defined				-2,09

CBU_1465	**hypothetical protein CBU_1465**	**-**	**COG2947S**	**-2,23**		**-2,46**
CBU_1468	**hypothetical protein CBU_1468**	**-**	**COG3164S**	**-2,10**		**-2,06**
CBU_1469	**rod shape-determining protein**	***mreD***	**COG2891M**	**-2,24**		**-2,38**
CBU_1470	**rod shape-determining protein**	***mreC***	**COG1792M**			**-2,24**
CBU_1515	**hypothetical protein CBU_1515**	**-**	**-**			**-2,11**
CBU_1611	**not defined**	**-**	**not defined**		**-2,02**	
CBU_1615	**Hypothetical protein (pseudo gene)**	**-**	**not defined**		**-2,13**	
CBU_1636	**hypothetical protein CBU_1636**	**-**	**-**	**2,01**		
CBU_1661	**lipopolysaccharide heptosyltransferase II**	***rfaF***	**COG0859M**		**2,02**	
CBU_1686	**hypothetical protein CBU_1686**	**-**	**-**	**2,44**		
CBU_1688	**deoxycytidine triphosphate deaminase**	***dcd***	**COG0717F**	**2,05**		
CBU_1705	**hypothetical protein CBU_1705**	**-**	**-**			**-2,20**
CBU_1708	**superoxide dismutase**	***sodB***	**COG0605P**			**-2,04**
CBU_1709	**dihydrodipicolinate reductase**	***dapB***	**COG0289E**			**-2,43**
CBU_1710	**hypothetical protein CBU_1710**	**-**	**-**			**-2,02**
CBU_1712	**GIY-YIG catalytic domain protein**	**-**	**COG2827L**			**-2,13**
CBU_1715	**glycine cleavage system H protein**	***gcvH***	**COG0509E**			**-2,03**
CBU_1720	**aconitate hydratase**	***acnA***	**COG1048C**			**-2,14**

CBU_1765	**hypothetical protein CBU_1765**	**-**	**-**		**2,11**		
CBU_1813	**hypothetical protein CBU_1813**	**-**	**-**	**2,00**			
CBU_1828	**hypothetical protein CBU_1828**	**-**	**COG1565S**		**-2,05**		
CBU_1838	**dTDP-4-dehydrorhamnose 3,5-epimerase**	***rfbC***	**COG1898M**	**2,31**			
CBU_1846	**not defined**	**-**	**not defined**	**2,21**	**2,08**		**2,16**
CBU_1850	**hypothetical protein CBU_1850**	**-**	**-**				**2,02**
CBU_1857	**oligopeptide transport ATP-binding protein**	***oppD***	**COG4172R**				**2,10**
CBU_1858	**oligopeptide transport system permease protein**	***oppC***	**COG1173EP**				**2,07**
CBU_1859	**oligopeptide ABC transporter, permease protein**	***oppB***	**COG0601EP**				**2,30**
CBU_1903	**cell division protein**	***ftsY***	**COG0552U**				**-2,02**
CBU_1920	**inner membrane protein oxaA**	***yidC***	**COG0706U**			**-2,04**	
CBU_1921	**hypothetical protein CBU_1921**	**-**	**-**			**-2,16**	
CBU_1922	**tRNA modification GTPase TrmE**	***trmE***	**COG0486R**			**-2,22**	
CBU_1924	**tRNA uridine 5-carboxymethylaminomethyl modification ...**	***gidA***	**COG0445D**			**-2,11**	
CBU_1925	**methyltransferase GidB**	***gidB***	**COG0357M**	**-2,10**		**-3,04**	
CBU_1927	**chromosome partitioning protein**	***parB***	**COG1475K**	**-2,05**		**-2,68**	
CBU_1928	**acyl-CoA synthetase**	**-**	**COG0204I. COG0318IQ. COG1020Q**			**-2,18**	

CBU_1947	**UDP-N-acetylglucosamine pyrophosphorylase**	***glmU***	**COG1207M**		**2,00**		
CBU_1964	**peptide release factor-glutamine N5-methyltransferase**	***hemK***	**COG2890J**		**2,21**		
CBU_1970	**diaminopimelate epimerase**	***dapF***	**COG0253E**		**2,41**		
CBU_1971	**hypothetical protein CBU_1971**	**-**	**-**		**2,55**		
CBU_1973	**hypothetical cytosolic protein**	**-**	**-**		**3,40**	**2,26**	
CBU_1975	**phospholipase/carboxylesterase family protein**	**-**	**COG0400R**		**3,15**	**2,62**	
CBU_1976	**nucleotidyl transferase family protein**	**-**	**COG1208MJ**		**3,60**	**2,92**	**2,07**
CBU_1977	**phosphotransferase enzyme family protein**	**-**	**COG3178R**		**2,14**	**2,06**	
CBU_1980	**putative peptidyl-prolyl cis-trans isomerase surA**	**-**	**COG0760O**		**2,18**		
CBU_1981	**4-hydroxythreonine-4-phosphate dehydrogenase**	***pdxA***	**COG1995H**		**2,26**		
CBU_1991	**RelB**	***relB***	**COG2161D**		**2,13**	**2,04**	
CBU_2016	**hypothetical cytosolic protein**	**-**	**-**			**2,29**	
CBU_2024	**putative cystathionine beta-synthase**	**-**	**COG0031E**			**2,12**	
CBU_2026	**Hypothetical exported protein (pseudo gene)**	**-**	**not defined**			**2,93**	
CBU_2030	**S-adenosylmethionine synthetase**	***metK***	**COG0192H**			**2,06**	
CBU_2069	**thymidine kinase**	**-**	**COG1435F**	**2,11**			

Table 1 : List of genes differentially expressed upon different short temperature stresses

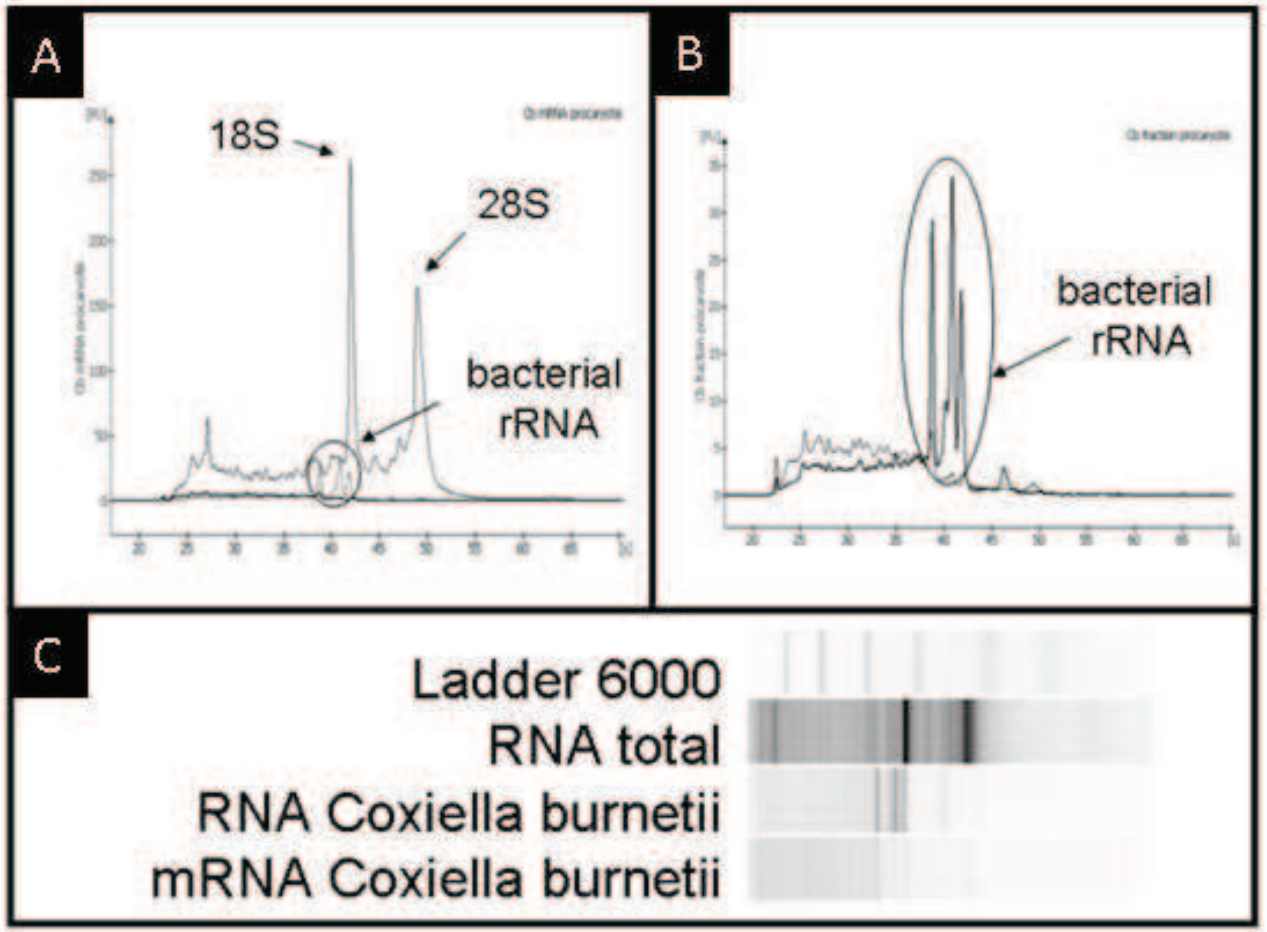

A
18S
28S
bacterial
rRNA
B
bacterial
rRNA
C
Ladder 6000
RNA total
RNA Coxiella burnetii
mRNA Coxiella burnetii

Figure 1

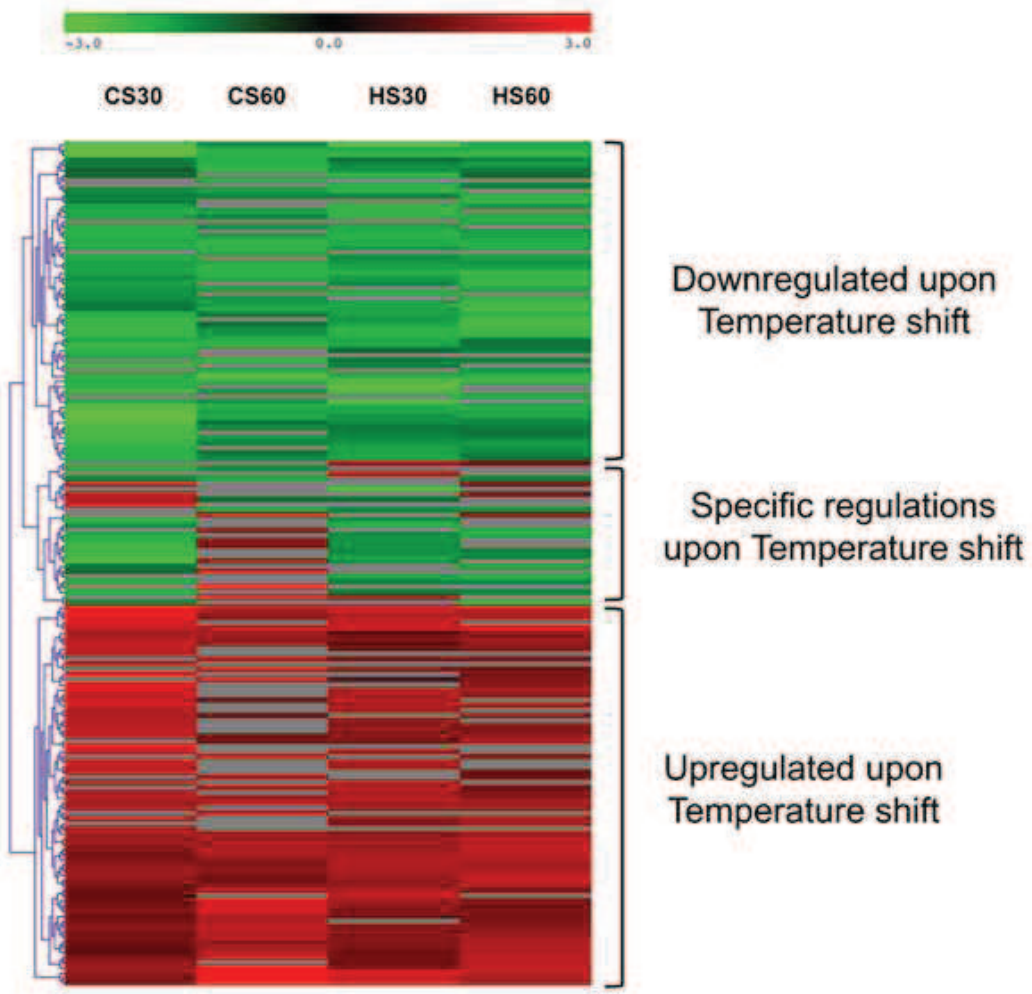

-3.0
0.0
3.0
CS30
CS60
HS30
HS60
Downregulated upon Temperature shift
Specific regulations upon Temperature shift
Upregulated upon Temperature shift

Figure 2

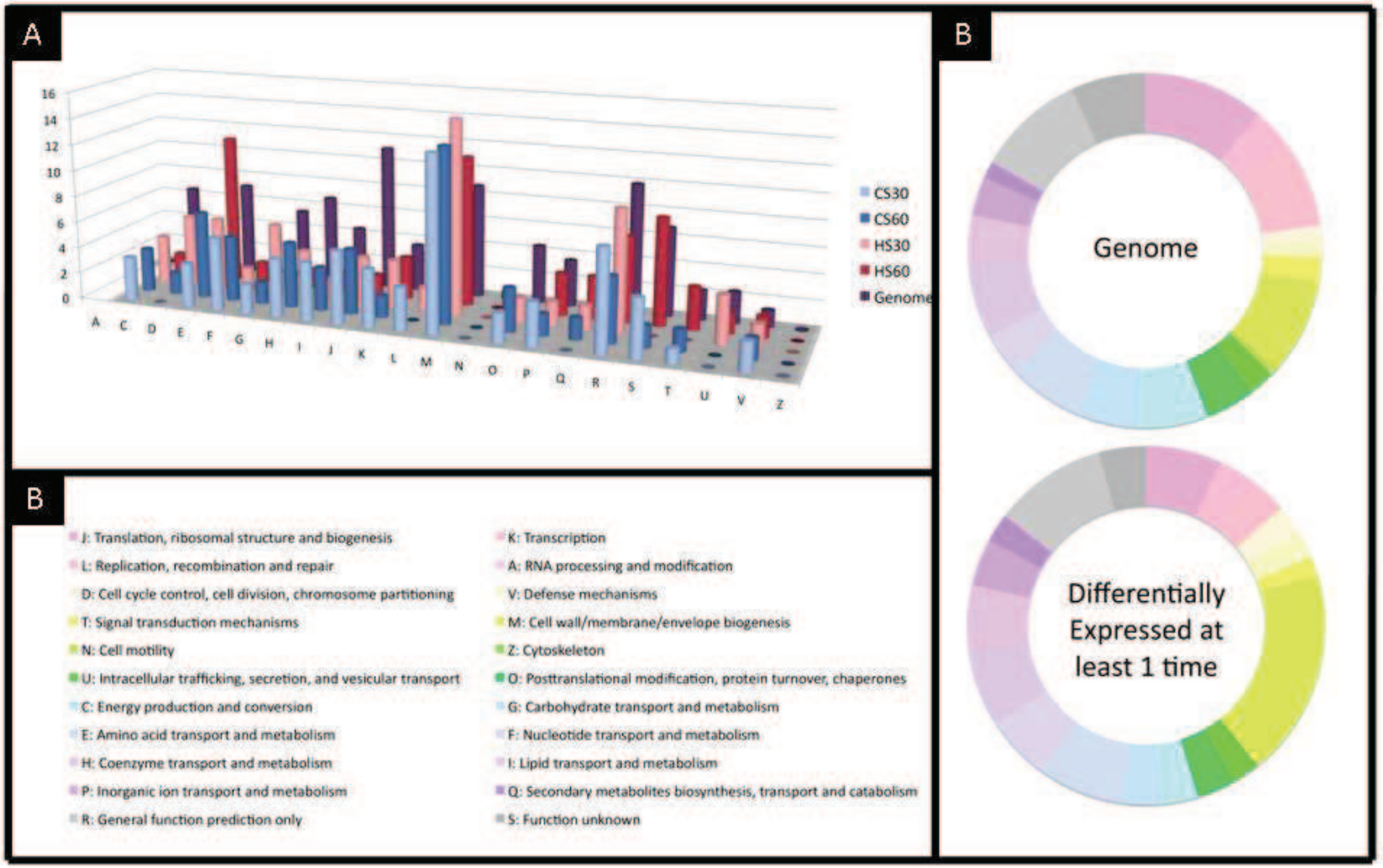

A
16
14
12
10
8
6
4
2
0
A C D E F G H I J K L M N O P Q R S T U V Z
CS30
CS60
HS30
HS60
Genome
B
J: Translation, ribosomal structure and biogenesis
K: Transcription
L: Replication, recombination and repair
A: RNA processing and modification
D: Cell cycle control, cell division, chromosome partitioning
V: Defense mechanisms
T: Signal transduction mechanisms
M: Cell wall/membrane/envelope biogenesis
N: Cell motility
Z: Cytoskeleton
U: Intracellular trafficking, secretion, and vesicular transport
O: Posttranslational modification, protein turnover, chaperones
C: Energy production and conversion
G: Carbohydrate transport and metabolism
E: Amino acid transport and metabolism
F: Nucleotide transport and metabolism
H: Coenzyme transport and metabolism
I: Lipid transport and metabolism
P: Inorganic ion transport and metabolism
Q: Secondary metabolites biosynthesis, transport and catabolism
R: General function prediction only
S: Function unknown
B
Genome
Differentially Expressed at least 1 time

Figure 3

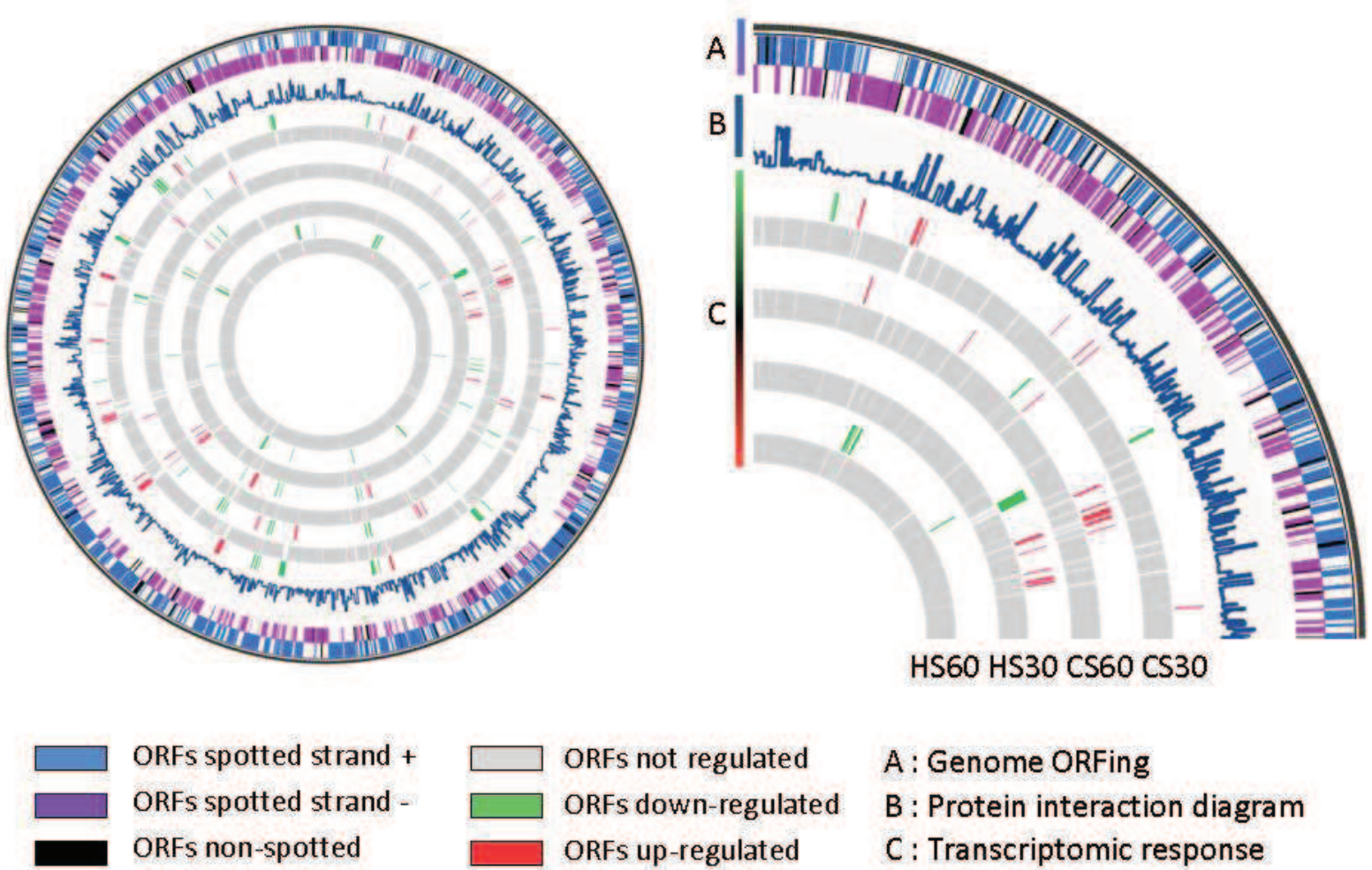

A
B
C
HS60 HS30 CS60 CS30
ORFs spotted strand +
ORFs spotted strand -
ORFs non-spotted
ORFs not regulated
ORFs down-regulated
ORFs up-regulated
A : Genome ORFing
B : Protein interaction diagram
C : Transcriptomic response

Figure 4

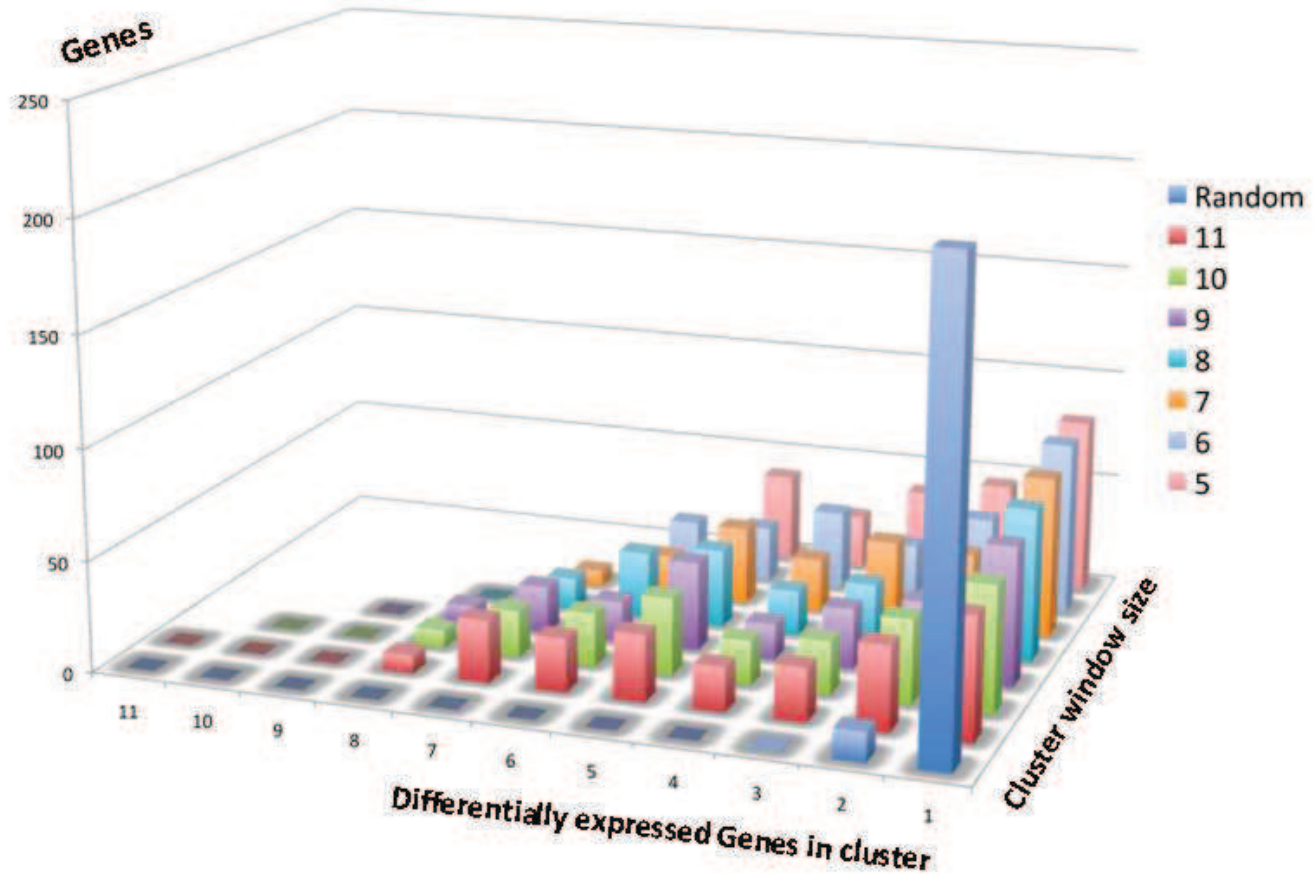

Genes
250
200
150
100
50
0
11
10
9
8
7
6
5
4
3
2
1
Differentially expressed Genes in cluster
Cluster window size
Random
11
10
9
8
7
6
5

Figure 5

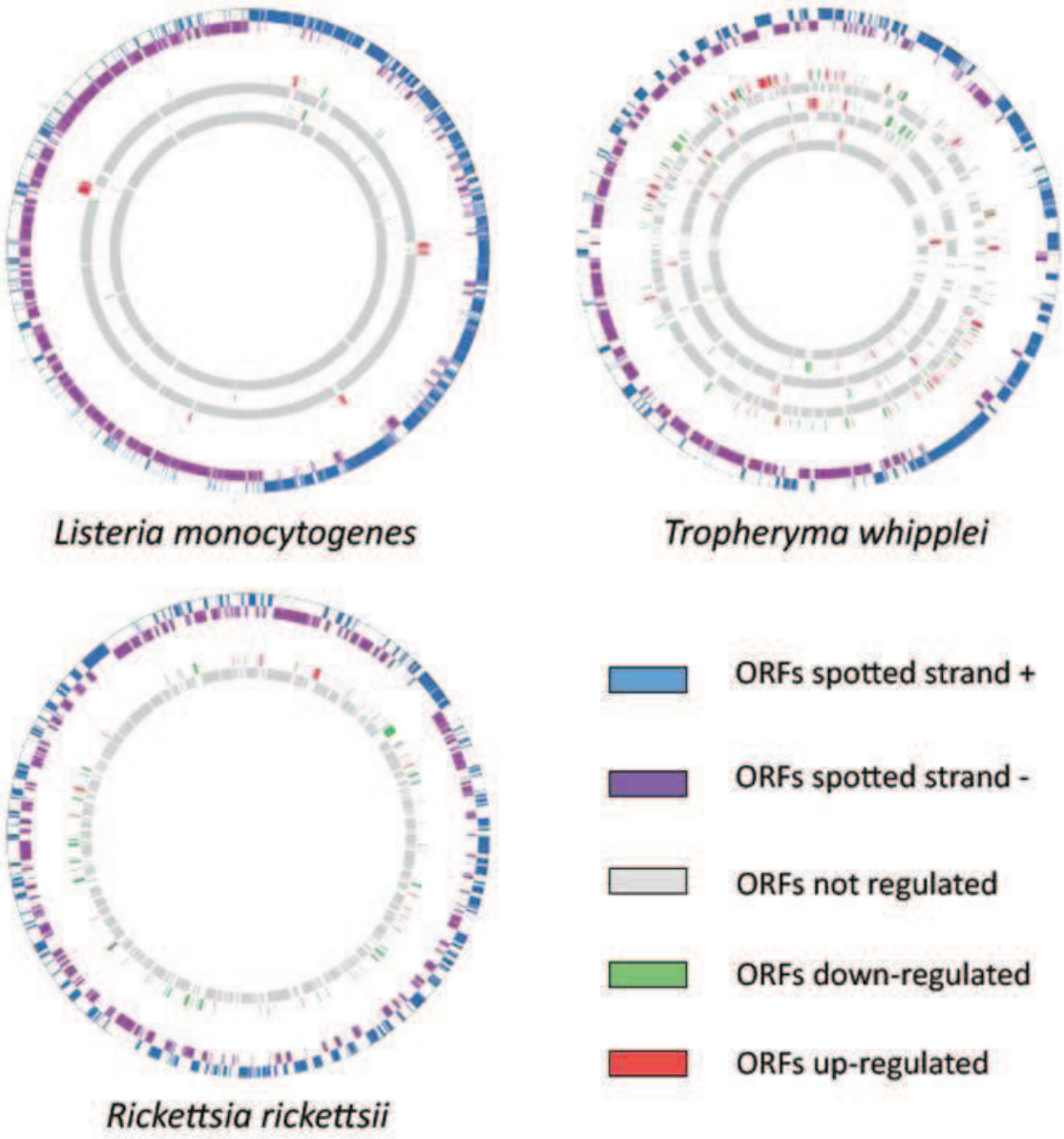
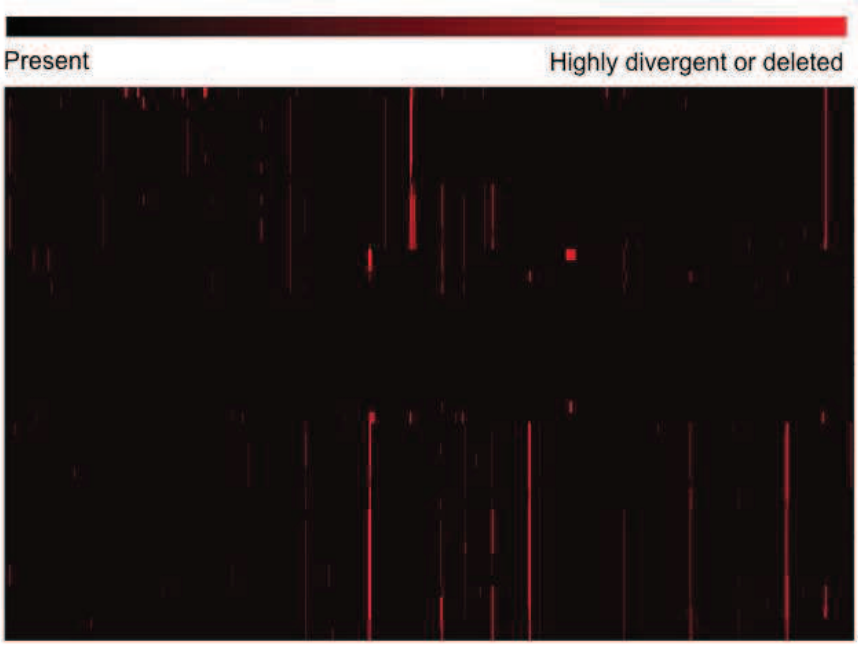

Listeria monocytogenes
Tropheryma whipplei
Rickettsia rickettsii
ORFs spotted strand +
ORFs spotted strand -
ORFs not regulated
ORFs down-regulated
ORFs up-regulated

Figure 6

Chapitre III

Genomotyping of 52 isolates of *Coxiella burnetii* reveals high homogeneity of gene content within isolates associates with ticks



Commentaire

La fièvre Q est une maladie ubiquitaire [32], bien qu'à ce jour aucun cas n'a été diagnostiqué en Nouvelle-Zélande. Les manifestations cliniques de même que les résultats biologiques ne sont pas spécifiques au cours de la fièvre Q, ce qui rend le diagnostic souvent difficile et dépendant du fait que la fièvre Q soit évoquée de façon systématique. La mise en évidence des cas dépend donc de l'intérêt du clinicien et de la présence d'un laboratoire de référence pour la fièvre Q [7].

L'analyse de la séquence de l'ARNr 16S a mis en évidence une forte homogénéité entre les souches de *C. burnetii* [50]. Par contre, les analyses par RFLP [51,52] ou PCR-RFLP [53,54] ont permis de mettre en évidence une diversité entre différentes souches. L'étude la plus complète de typage a été rapportée par Glazunova et *al.* [55], qui ont utilisé le Multi Spacer Typing sur environ 150 isolats de *C. burnetii*. Plus récemment, une étude de génomique comparative par puces à ADN a été réalisée sur une collection de 24 souches dans le but de corréler le répertoire génétique et la source de l'organisme [56].

Récemment, une longue controverse sur la virulence des différentes souches de *C. burnetii* a été résolue [55,57–59]. En effet, il semblerait que tous les génotypes de souches peuvent être isolés lors d'infections chroniques, laissant penser ces dernières seraient donc plus déterminées par les facteurs liés à l'hôte. En revanche, seuls certains génotypes ont été isolés lors de fièvres Q aiguës. Par ailleurs, il a été démontré que la souche de référence Nine Mile induisait une infection aiguë à plus faible inoculum que la souche Q212 (isolée à partir d'une infection chronique). Par conséquent, il existe des différences de virulence entre les souches qui pourraient être à l'origine des infections aiguës.

Les différentes méthodes de typage (Plasmide, Multi Spacer Typing et génomotypage) ont montré une capacité à discriminer ces souches [55] [56].

Actuellement, la fièvre Q se voit porter un regain d'intérêt du fait de sa réémergence dans différentes zones et spécialement en Europe avec l'immense épidémie qui fait rage aux Pays-Bas [36,40]. Ces récentes épidémies de fièvre Q ont relancé la question de la clonalité bactérienne. En effet, il est possible que ces épidémies soient liées à l'émission de clones hautement virulents. De nombreuses études de biologie moléculaire ont été effectuées lors de l'épidémie aux Pays-Bas et les résultats ont montré l'implication d'un seul génotype ou du moins une réduction de l'hétérogénéité [40].

Le rôle des tiques en tant que vecteur ou réservoir a été discuté [60–64]. Même si les grandes épidémies semblent être liées à l'exposition à des animaux, le rôle des arthropodes dans la transmission de la fièvre Q doit être considéré. Il a été montré que les tiques peuvent être infectés par *C. burnetii* lors repas sanguin, elles excrètent la bactérie *via* les selles, la salive et le fluide coxal et peuvent également la transmettre par voie transovarienne et transtadialle. Aujourd'hui, la transmission vectorielle de *C. burnetii* est considérée comme une voix mineure dans la dissémination de la fièvre Q [35].

Dans cette étude, nous avons génomotypé à l'aide de puces à ADN une collection de 52 isolats provenant de patients et d'animaux (mammifères, oiseaux et tiques) dont 2 isolats présentant le même génotype que le clone épidémique présumé de Pays-Bas en se basant sur la MST, dans le but de trouver des associations entre le répertoire génétique, la source d'isolement et les informations cliniques des différents isolats de *C. burnetii*.

Cet article nous a permis de mettre en évidence une congruence entre le génomotypage et le Multi Spacer Typing. Nous avons aussi augmenté le nombre

de génomotypes associés aux manifestations aiguës de la fièvre Q. De plus, nous avons identifié 4 gènes dont la délétion est significativement associée à la forme aiguë de la maladie. Enfin, nous avons pu mettre en évidence que les isolats provenant des humains présentaient significativement plus d'événements de délétion que les isolats provenant d'animaux et spécialement ceux de tiques dures qui ne présentent aucune différence avec la souche de référence Nine Mile.

Article 3

Genomotyping of 52 isolates of *Coxiella burnetii* reveals high homogeneity of gene content within isolates associates with ticks

Quentin LEROY[1], Fabrice ARMOUGOM[1], Pascal BARBRY[2], Didier RAOULT[1]

En préparation

1 - Unité des Rickettsies, IRD-CNRS UMR 6236, Faculté de Médecine, Marseille, France

2 – Institut de Pharmacologie Moléculaire et Cellulaire (IPMC), UMR 6079 CNRS-UNSA, Sophia Antipolis, France

Abstract

C. burnetii is a Gram-negative intracellular γ-proteobacteria that causes Q fever, a zoonotic disease. Q fever can manifest as an acute or chronic illness. Different typing methods have been previously developed to classify *C. burnetii* isolates in order to explore its pathogenicity. Here, we report a genomotyping method based on presence or absence of genes using microarray based on Nine Mile strain genome. The genomotyping method was then tested on 52 isolates obtained from different geographic areas, different hosts and isolated from patient with different clinical manifestations. The analysis reveals the presence of 10 genomotypes organized in 3 groups with a topology congruent with that of Multi Spacer Typing. We also found out 4 genomotypes especially associated with acute Q fever whereas all the genomotypes could be associated to chronic human infection. Serendipity, genomotyping reveals that hard ticks isolates including Nine Mile belong to the same genomotype.

Introduction

C. burnetii is a Gram-negative intracellular γ-proteobacteria that causes Q fever, a zoonotic disease with a worldwide distribution [1]. Q fever can manifest as an acute or chronic illness. Acute Q fever is typically a self-limiting febrile illness during which pneumonia or hepatitis can occur, whereas chronic Q fever is a severe illness where patients can present endocarditis, vascular infection, osteomyelitis and chronic hepatitis [1]. Aerosol is the major route of contamination with *C. burnetii*. *C. burnetii* displays antigenic variation in the lipo-polysaccharide (LPS) [2]. The phase I is highly infectious and corresponds to the natural phase found in animals, human and arthropod, whereas phase II that is not very infectious, presents a truncated LPS and can be obtained after several passages on cell culture or embryonated eggs [1]. *C. burnetii* genome was sequenced in 2003 and its size is around 2Mbp with a plasmid of around 38 kpb [3]. Recently 5 new isolates were sequenced [4].

Analysis of 16S rDNA gene sequencing has highlighted that *C. burnetii* strains isolated from a variety of geographical areas and various hosts display considerable genetic homogeneity [5]. Restriction fragment length polymorphism (RFLP) analysis of genomic DNA (gDNA) [6–8] and sequence and/or PCR-RFLP [9–12] of specific genes reveals genetic diversity between *C. burnetii* isolates. The most extensive survey of *C. burnetii* genetic diversity was reported by Glazunova et al. [13], who used multi spacer sequence typing to genotype around 150 *C. burnetii* isolates. More recently, a comparative genomic hybridization (CGH) analysis were performed on a collection a 24 strains of *C. burnetii* [14]. The *C. burnetii* genome sequence availability allows a rapid assessment of whole genome sequence variation by using comparative genome hybridization CGH on microarrays allowing correlating the genome repertoire and the source of the organisms.

A long controversy on the virulence of different isolates of *C. burnetii* has been resolved recently [13,15–18]. It appears that all type of strains can be isolated from chronic infections that are more determined by host factors that by bacterial factors. In contrast, only some strains have been isolated from acute infection and the prototype strain Nine Mile have been found to determine acute infection at lower inoculum than the strain Q212 found out in a chronic infection. Therefore, there is difference of strains virulence in causing acute infection that matches with the genotype determined by MST, genomotyping or plasmid typing.

Q fever is currently re-emerging in different areas in Europe with the gigantic outbreak of Q fever observed in Netherlands (causing both acute and chronic infections) [19] and in US military personal affected in Iraq [20]. *C. burnetii* outbreaks highlight the question of bacterial clonality that could be related to broadcasts of highly virulent clones. Alternatively, the apparent massive increase in cases of Q fever could be actually related to improved detection or increased risk of exposure to animal reservoir [20]. The outbreak now raging in Netherlands has been the subject of different molecular biology experiments including one that reveal the implication of a single genotype or at least a reduction of heterogeneity [19,21]. We had the opportunity to test this genotype using MST methods [13] and found that the putative clone responsive of the outbreak was identical to a strain isolated from an infected sheep vagina in Germany for over 10 years and several strains isolated from humans in France (unpublished data).

Even if the large outbreaks seem to be related to exposure of animal, the role of arthropods in Q fever transmission has to be considered. The role of ticks as vector and reservoirs has been discussed since 1937. Ticks may be infected by *C. burnetii* during feeding, excrete it *via* faeces, saliva and coxal fluid and

may transmit it transovarially and transstadially. The reference strain Nine Mile was isolated from a *Dermacentor andersoni* hard tick and Q fever was initially presumed to be a vector-borne disease [22–25]. Nowadays, ticks are rare vector for transmission of Q fever [26].

In this study, we compare 52 isolates of from patients and animals (mammals, bird and ticks) from our strain collection including 2 isolates presenting the same MST genotype than the putative epidemic clone from Netherland using DNA whole genome microarray to perform genomotyping to investigate associations of gene repertoire, source and clinical information of *C. burnetii.*

Materials and methods

C. burnetii isolated, cultivation and purification

Isolate name, geographical/sample origin, plasmid type, clinical disease are listed in Table 1. *C. burnetii* grown at 35°C on L929 cells with MEM medium (GIBCO, Invitrogen, Cergy-Pontoise, France) supplemented with 4% SVF (GIBCO) and 1% of L-Glutamine (GIBCO). Monolayers cells and supernatant from 3 175cm² flask were harvested and incubated with 1% of Trypsine (GIBCO) for 1 hour at 37°C. Released bacteria were purified from L929 cell debris by differential centrifugation. Purified bacteria were resuspended in 400µl of PBS and stored at -80°C.

gDNA extraction and amplification

200 µl of purified bacteria were incubated for 30 min at 70°C with 200 µl of AL lysis buffer (Qiagen, Courtaboeuf, France) and 20 µl of Proteinase K (Qiagen). gDNA were extracted and purified using QiaAmp DNA mini kit as recommended by the manufacturer (Qiagen). gDNA purity and concentration was checked using the NanoDrop (Thermo, Wingmilton, USA). 10 ng of gDNA were amplified by the processive polymerase phi29 using the GenomPhi illustrator V2 kit (GE HealthCare, Lifescience, Orsay, France). This strategy was previously described for CGH experiments [14].

gDNA labelling and microarray experiments

The amplified gDNA were labeled with the Bioprime CGH Labelling kit (Invitrogen) using d-CTP Cy3/5 fluorochromes (GE HealthCare Lifescience) as recommended by the manufacturer. Labeled amplified gDNA were purified using Pure Link PCR purification columns (Invitrogen) and the level of fluorochromes incorporation was quantified using the NanoDrop (Thermo Scientific). Hybridizations were carried out using two samples of labeled amplified gDNA (150 pmol of each) that were labeled with Cy3 or Cy5 d-CTP. The pooled samples were hybridized using the GE hybridization kit (Agilent

Technologies) as recommended by the manufacturer. The mixture was applied on a Surhyb 1 array (Agilent Technologies) and hybridized on the *Coxiella burnetii* array using the Agilent hybridization chamber (Agilent Technologies). Microarrays were hybridized for 17h at 62°C in a rotative oven. Microarrays were washed using the GE washing buffers (Agilent Technologies) with 5 min of Wash buffer 1 at room temperature followed by 1 min of Wash buffer 2 at 37°C. Microarrays were dried using a bath of acetonitrile (VWR, Fontenay sous Bois, France). The microarrays were scanned using the microarray scanner C (Agilent Technologies) using XDR at 5µm resolution.

***Coxiella burnetii* whole genome microarray construction**

OligoArray 2.0 [27,28]was used to calculate probes from 2016 CDS extracted from NC_002971.gb Genbank sequence file corresponding to that of Nine Mile reference strain [3]. OligoArray 2.0 integrates BLAST analysis against a non-redundant set of sequences and probe secondary structure analyses [29]. Oligonucleotide calculation parameters were set as follows: oligo length from 50 to 52mers; GC percentage from 35 to 55%; melting temperature from 82 to 86°C. OligoArray 2.0 selected probes with the lowest cross-hybridization, the absence of secondary structure and balanced the set of probes in terms of melting temperature. Oligonucleotides containing five consecutive A, C, G or T's were discarded. Following probes design, 1990 probes corresponding to 1990 distinct CDS where selected for synthesis. A total of 100 µmoles of each probe were ordered from Sigma–Proligo (Paris, FRANCE) as a 5′ amino modified oligonucleotide. Oligonucleotide stocks were aliquoted for use in microarray fabrication. Oligonucleotides were diluted to a final concentration of 35–50 µM in 35% dimethyl sulfoxide (DMSO), 100 mM potassium phosphate (pH 8.0). Coxiella burnetti - 2k microarrays were printed with a ChipWriterProarrayer (Bio-Rad Hercules, CA) on commercial HydroGel slides (Schott, Mainz, Germany), and processed according to the manufacturer's

instructions. Our microarray is spotted in quadriplicate and contains 1990 different probes genes that correspond to ca. 98.7% of the coding sequences.

Analysis of microarray data

All microarray results have been deposited in the GEO database (http://www.ncbi.nlm.nih.gov/geo/) under GEO series accession number (A faire). The signal intensity and local background were measured for each spot using the array pictures with Feature Extractor software (Agilent Technologies). Data filtering normalizations were obtained by using processing signal raw from obtained after data extraction from Feature Extractor. We used to mean of our four replicated per probes to performed M-A plot. Using M-A plots, we deduce a naïve cut-off [30] to obtained genes that are putatively lost compared or highly divergent to our reference. Matrix for clusterization of our data was performed using 0 for gene conserved and 1 for gene putatively lost or highly divergent. Clustering analyses were performed using Tmev [31,32]. We used hierarchical clustering to generate the dendograme with Euclidean distance and complete linkage for respectively distance metric calculation and linkage method.

Genomotyping and statistical analysis.

To performed genomotyping we identify the putative single events of mutation. Matrix for clusterization of our data was performed using 0 for genomic content conserved and 1 for putative events of mutation. Clustering analyses were performed using Tmev [31,32]. We used hierarchical clustering to generate the dendograme with Euclidean distance and complete linkage for respectively distance metric calculation and linkage method. The statistical analysis was performed using GraphPad Prism5 (GraphPad Software, Inc). The principal component analysis was performed using Comprehensive Meta-analysis software (Biostat, Englewood NJ).

Results

CGH experiments

The CGH experiment using a whole genome microarray was performed to genomotype 52 isolates of *C. burnetii* to detect deleted genes compared to the reference strain Nine Mile. The isolation and clinical information from the collection of isolates are listed in table 1. To confirm the gene losses we compared results to the previous study of CGH for *C. burnetii* [14]. We found comparable results with our deleted genes set and the set obtained by the previous CGH study. Strains used in both studies (HzS and S217), present similar gene content. Given the putative gene loss deduced from the 52 hybridization data, it appeared that *C. burnetii* genomic content is highly conserved across the 52 tested isolates (Table 1). The chromosomal deletion associated with the phase II conversion [33] was found in only two isolates (HzR and Luga) and will not be included in our study. Comparative analysis showed that relative to the NMI strain, the percentage of deleted or highly divergent ORFs ranges from 0 to 2.5 % (S217). The heat map visualization of genomic variations showed that differences are spread across the genome (Supplementary data). Only 161 genes from the NMI isolate were predicted to be absent or highly divergent in at least one tested strain (Supplementary data). The clustering analysis of gene putatively deleted at least in 1 isolate show 3 distinct clusters 1, 2 and 3 (Supplementary data). The clusters 1 and 2 contained genes with high deletion frequency whereas the cluster 3 is mostly composed of 95 genes with globally low deletion frequency.

Genomotyping

To perform genomotyping, we first identify the deleted or highly divergent genes and assigned them single chromosomal mutation events (supplementary table 2) based on the methods used for the previous study of Beare et *al.* [14]. We found as shown in Figure 1A that isolates are organized in

two major (A and B) and one minor (C) groups that contained respectively 21, 30 and 1 isolates (Figure 1). The groups A, and B are constituted respectively of 3 (A1 to A3), 6 (B1 to B6), distinct genomotypes. Within genomotypes, low variability of genomic content is observable. However, only few isolates have identical gene content and small divergences occur within genomotypes. We found also that the group A is associated to deleted gene cluster 1. The cluster 2 and 3 could be found associated with the three groups A, B and C.

MST-typing and genomotyping

We compared the genomotyping to MST genotyping previously described [13] sharing 3 groups presenting a similar topology (Figure 1B). The MST genotypes 1 to 10 are included within the group 1. The genotypes 21 are included within the group 2 and the other MST genotypes are associated with the group 2. Genomotype groups A, C and B includes MST-genotypes 1 to 8, MST genotype 21 and the other MST-genotypes. We found only three exceptions with the isolates CB76, CB93 and CB94 that have not been associated within the expected genomotype groups. Despite the figure 1B highlights low divergence between genomotyping and MST-genotyping, the two different methods reveal congruence in isolates clusterization.

Correlation between gene content and physiopathology

We wanted to know if the gene content of different strains could be correlated with physiopathological and geographical information listed in table 1. We try to associate genes with Q fever acute clinical form. We found that 4 clusters were associated to acute infection (A2, A3, B4 and B5). We consider that the 4 clusters were the genomotypes that could cause acute infection. In this regard, we focus on genes that are specially deleted in acute isolates and their clusters and present for chronic clusters and *vice versa*. We found that 4 deleted ORFs were significantly associated to the acute infection and their clusters

(Table 2). These genes are annoted as hypothetical. We found also that the two isolates with the same genotype than the putative epidemic clone (CB74 and Z3055) present few differences of gene repertoire and belong to different genomotypes (B3 and B2). Moreover, human isolates seems to have less genes than animal isolates (Figure 2A). Furthermore, arthropod isolates present less deleted genes and specially the hard ticks (Luga, Derma, 5116 and NM) which do not present deleted gene (genomotype B1). We also found that isolates associated with the plasmid QpH1 has less deleted genes compared to the plasmids QpRS and QpDV (Figure 2B). Animal isolates were also principally associated with QpH1 plasmid type (Table 1). We performed principal component analysis (PCA) to observe association between gene absences or genomotypes and clinical or geographical information. We could not find any obvious associations with gene absence, but confirm as previously mentioned that B1 genomotype is associated to arthropods and especially hard ticks (Figure 2C).

Discussion

In this study, we examined the genomic content of 52 isolates of *C. burnetii* compared to the reference strain NMI, using a global genomic approach based on comparative genomic hybridization by whole genome microarray. A previous CGH analysis of *C. burnetii* was performed on 23 different isolates and two antigenic variant of NMI. Our collection has 3 strains in common with that used by Beare et *al*. (NMI, S217 and HzS) that we used as positive control for our work. We found high homologies between our results and that of Beare and al. [14]. In this regard, we consider that we obtained robust and confirmed data to performed genomotyping with our microarray results. Several typing methods had been developed for the causative agent of the Q fever [5,7,8,10,12–14]. Glazunova et *al.* [13] and Beare et *al.* [14] have shown that the different methods of typing were globally congruent. In our study, we have compared whole-genome typing with MST methods. The comparison shows a similar clusterization of isolates within 3 groups with significantly divergent gene content (Figure 1B). In this regards, the different isolates of *C. burnetii* present convergent evolutions that are independent of geographic origin and clinical context as previously proposed by Glazunova et *al.* [13] and Beare et *al.* [14].

Several animal models have shown that both the inoculum size used and the strain influence the presence and manifestations of acute pneumonia during Q fever [17,18,34]. However, there is no evidence that isolates from chronic and acute human infections differ when large collections are screened by different methods of typing [13]. The preliminary work based on MST-typing have shown that acute Q fever was induced by isolates belonging to genotype MST 1, 2, 4, 16, 18 and that the plasmid QpDV was highly associated with acute Q fever. Isolates from chronic infections were associated with all MST-

genotypes and all of the plasmid type. In our study we found that only genomotypes A2, A3, B4 and B5 contained isolates from acute infections. Based on these associations, we have found 4 deleted or highly divergent ORFs with unknown function that were significantly associated with acute Q fever. Furthermore the previous microarray genomic analysis showed that isolates with acute infections had comparable gene repertoire to that of genomotype B1 [14]. As no genomotypes is specifically associated with chronic Q fever, we confirmed that all isolates could be involved in chronic infections as previously proposed [13,16]. We described 4 more genomotypes associated with acute infections, compared to the previous study. Based on our results, the previous typing methods of isolates collection and on studies of acute Q fever animal models we proposed the diagram of Q fever infections (figure 3).

Surprisingly, the comparative genomic analysis of our isolates collection has revealed that animal isolates (and especially arthropods) has globally less deleted genes than human isolates (figure 3A). All isolates from hard tick present identical gene contents with Nine Mile I. Beare et *al.* [14] had previously found that 2 isolates associated with hard ticks have identical gene content with NMI (Dugway 5G61-63 and BDT 1). These isolates were isolated from tick species. *C. burnetii* has been identified in many species of ticks [26].

A bias in sampling exists in this study. Whereas chronic disease represents 20 times less patients than acute disease, most of isolates the human isolates were from chronic disease patients, and the isolates from acute infections were mainly obtained from France. However, our isolates collection contains 2 isolates presenting the identical MST-genotype than the putative epidemic clone that coming from a patient suffering of a chronic Q fever in Marseille (CB74) and from a placenta of goat in Germany (Z3055). The

2 isolated reveals comparable but different gene repertoires and are associated to different genomotypes that belong to the genomotype group B.

A major limitation of our study is that we only used the genome of the reference strain Nine Mile [3]. It was the only available at the beginning of our study and addition of sequence from further strains may contribute to a better understanding of *C. burnetii* cycle and Q fever pathophysiology. However we found here for the first time that isolates from hard ticks have the same gene content and probably the same origin. Moreover we found that loss of 2 putative genes associated with virulence fuelling the hypothesis that bacterial pathogenicity is driven more by gene loss than gene gain.

Reference List

1. Maurin M, Raoult D (1999) Q fever. Clin Microbiol Rev 12: 518-553.

2. Hackstadt T, Peacock MG, Hitchcock PJ, Cole RL (1985) Lipopolysaccharide variation in Coxiella burnetti: intrastrain heterogeneity in structure and antigenicity. Infect Immun 48: 359-365.

3. Seshadri R, Paulsen IT, Eisen JA, Read TD, Nelson KE, Nelson WC, Ward NL, Tettelin H, Davidsen TM, Beanan MJ, Deboy RT, Daugherty SC, Brinkac LM, Madupu R, Dodson RJ, Khouri HM, Lee KH, Carty HA, Scanlan D, Heinzen RA, Thompson HA, Samuel JE, Fraser CM, Heidelberg JF (2003) Complete genome sequence of the Q-fever pathogen *Coxiella burnetii*. Proc Natl Acad Sci U S A 100: 5455-5460. 10.1073/pnas.0931379100 [doi];0931379100 [pii].

4. Beare PA, Unsworth N, Andoh M, Voth DE, Omsland A, Gilk SD, Williams KP, Sobral BW, Kupko JJ, III, Porcella SF, Samuel JE, Heinzen RA (2009) Comparative genomics reveal extensive transposon-mediated genomic plasticity and diversity among potential effector proteins within the genus Coxiella. Infect Immun 77: 642-656. IAI.01141-08 [pii];10.1128/IAI.01141-08 [doi].

5. Stein A, Raoult D (1993) Lack of pathotype specific gene in human Coxiella burnetii isolates. Microb Pathog 15: 177-185. S0882-4010(83)71068-5 [pii];10.1006/mpat.1993.1068 [doi].

6. Heinzen R, Stiegler GL, Whiting LL, Schmitt SA, Mallavia LP, Frazier ME (1990) Use of pulsed field gel electrophoresis to differentiate Coxiella burnetii strains. Ann N Y Acad Sci 590: 504-513.

7. Hendrix LR, Samuel JE, Mallavia LP (1991) Differentiation of Coxiella burnetii isolates by analysis of restriction-endonuclease-digested DNA separated by SDS-PAGE. J Gen Microbiol 137: 269-276.

8. Jager C, Willems H, Thiele D, Baljer G (1998) Molecular characterization of Coxiella burnetii isolates. Epidemiol Infect 120: 157-164.

9. Zhang GQ, To H, Yamaguchi T, Fukushi H, Hirai K (1997) Differentiation of Coxiella burnetii by sequence analysis of the gene (com1) encoding a 27-kDa outer membrane protein. Microbiol Immunol 41: 871-877.

10. Nguyen SV, Hirai K (1999) Differentiation of Coxiella burnetii isolates by sequence determination and PCR-restriction fragment length polymorphism analysis of isocitrate dehydrogenase gene. FEMS Microbiol Lett 180: 249-254. S0378-1097(99)00495-4 [pii].

11. Sekeyova Z, Roux V, Raoult D (1999) Intraspecies diversity of Coxiella burnetii as revealed by com1 and mucZ sequence comparison. FEMS Microbiol Lett 180: 61-67. S0378-1097(99)00462-0 [pii].

12. Andoh M, Nagaoka H, Yamaguchi T, Fukushi H, Hirai K (2004) Comparison of Japanese isolates of Coxiella burnetii by PCR-RFLP and sequence analysis. Microbiol Immunol 48: 971-975. JST.JSTAGE/mandi/48.971 [pii].

13. Glazunova O, Roux V, Freylikman O, Sekeyova Z, Fournous G, Tyczka J, Tokarevich N, Kovacava E, Marrie TJ, Raoult D (2005) Coxiella burnetii genotyping. Emerg Infect Dis 11: 1211-1217.

14. Beare PA, Samuel JE, Howe D, Virtaneva K, Porcella SF, Heinzen RA (2006) Genetic diversity of the Q fever agent, *Coxiella burnetii*, assessed by microarray-based whole-genome comparisons. J Bacteriol 188: 2309-2324. 188/7/2309 [pii];10.1128/JB.188.7.2309-2324.2006 [doi].

15. Stein A, Lepidi H, Mege JL, Marrie TJ, Raoult D (2000) Repeated pregnancies in BALB/c mice infected with Coxiella burnetii cause disseminated infection, resulting in stillbirth and endocarditis. J Infect Dis 181: 188-194. JID990157 [pii];10.1086/315166 [doi].

16. Raoult D, Marrie T, Mege J (2005) Natural history and pathophysiology of Q fever. Lancet Infect Dis 5: 219-226. S1473-3099(05)70052-9 [pii];10.1016/S1473-3099(05)70052-9 [doi].

17. Russell-Lodrigue KE, Zhang GQ, McMurray DN, Samuel JE (2006) Clinical and pathologic changes in a guinea pig aerosol challenge model of acute Q fever. Infect Immun 74: 6085-6091. 74/11/6085 [pii];10.1128/IAI.00763-06 [doi].

18. Russell-Lodrigue KE, Andoh M, Poels MW, Shive HR, Weeks BR, Zhang GQ, Tersteeg C, Masegi T, Hotta A, Yamaguchi T, Fukushi H, Hirai K, McMurray DN, Samuel JE (2009) Coxiella burnetii isolates cause genogroup-specific virulence in mouse and guinea pig models of acute Q fever. Infect Immun 77: 5640-5650. IAI.00851-09 [pii];10.1128/IAI.00851-09 [doi].

19. Enserink M (2010) Infectious diseases. Questions abound in Q-fever explosion in the Netherlands. Science 327: 266-267. 327/5963/266-a [pii] ; 10.1126 /science. 327.5963.266-a [doi].

20. Raoult D (2009) Reemergence of Q fever after 11 September 2001. Clin Infect Dis 48: 558-559. 10.1086/596706 [doi];10.1086/596706 [pii].

21. Klaassen CH, Nabuurs-Franssen MH, Tilburg JJ, Hamans MA, Horrevorts AM (2009) Multigenotype Q fever outbreak, the Netherlands. Emerg Infect Dis 15: 613-614.

22. Weyer F (1953) Die Bezichungen des Q-Fieber Erregers (Rickettsia burneti) zu Arthropoden. Z Tropenmed Parasitol 4(3): 344-382.

23. Philip CB (1948) Observations on experimental Q fever. J Parasitol 34: 457-464.

24. Balashov, Y. S. and Daiter, A. B. (1973) Blood-feeding artropods and rickettsiae [In Russian]. Leningrad: Nauka (Russian Academy of Sciences). 251 p.

25. Smith DJW, Brown HE, Derrick EH (1939) Further series of laboratoriy infections with the Rickettsia of "Q" fever. Med J Australia 1: 13-20.

26. Mediannikov O, Fenollar F, Socolovschi C, Diatta G, Bassene H, Molez JF, Sokhna C, Trape JF, Raoult D (2010) Coxiella burnetii in humans and ticks in rural Senegal. PLoS Negl Trop Dis 4: e654. 10.1371/journal.pntd.0000654 [doi].

27. Rouillard JM, Herbert CJ, Zuker M (2002) OligoArray: genome-scale oligonucleotide design for microarrays. Bioinformatics 18: 486-487.

28. Rouillard JM, Zuker M, Gulari E (2003) OligoArray 2.0: design of oligonucleotide probes for DNA microarrays using a thermodynamic approach. Nucleic Acids Res 31: 3057-3062.

29. Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31: 3406-3415.

30. Carter B, Wu G, Woodward MJ, Anjum MF (2008) A process for analysis of microarray comparative genomics hybridisation studies for bacterial genomes. BMC Genomics 9: 53. 1471-2164-9-53 [pii];10.1186/1471-2164-9-53 [doi].

31. Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M, Sturn A, Snuffin M, Rezantsev A, Popov D, Ryltsov A, Kostukovich E, Borisovsky I, Liu Z, Vinsavich A, Trush V, Quackenbush J (2003) TM4: a free, open-source system for microarray data management and analysis. Biotechniques 34: 374-378.

32. Saeed AI, Bhagabati NK, Braisted JC, Liang W, Sharov V, Howe EA, Li J, Thiagarajan M, White JA, Quackenbush J (2006) TM4 microarray software suite. Methods Enzymol 411: 134-193. S0076-6879(06)11009-5 [pii];10.1016/S0076-6879(06)11009-5 [doi].

33. Hoover TA, Culp DW, Vodkin MH, Williams JC, Thompson HA (2002) Chromosomal DNA deletions explain phenotypic characteristics of two antigenic variants, phase II and RSA 514 (crazy), of the *Coxiella burnetii* Nine Mile strain. Infect Immun 70: 6726-6733.

34. Stein A, Louveau C, Lepidi H, Ricci F, Baylac P, Davoust B, Raoult D (2004) Q fever pneumonia: virulence of *Coxiella burnetii* pathovars in a murine model of aerosol infection. Infect Immun In press.

Figure Legends

Table 1: *C. burnetii* isolates used in this study

Table 2: ORFs associated with acute infections

Figure 1: Typing of the collection of isolates

(A) Genomic content clusterization of the isolates based on event of mutations that allow determining the different genomotypes. (B) Comparison of topology for genomotyping and MST-typing.

Figure 2: Association with gene repertoires and information

(A) The figure represents the association of the source of isolation and the number of deleted genes. (B) The figure represents the association of the plasmid type and the number of deleted genes. (C) The figure represents the PCA analysis of source and genomotyping. The blue circle represents the highlight associations.

Figure 3: Bacterial factors involved in Q fever

Supplementary data

Supplementary table 1: Representation of genomic content of the 52 isolates

Supplementary figure 1: Clusterization of deleted genes at least 1 time in the collection of isolates

Supplementary figure 2: Frequency of deletion

Name	Symptomes	Source	Country	MST	GT	Isolated from	Plasmid
5116	-	Hard Tick	Slovakia	ND	B1	Tick	ND
CB40	Chronic	Human	Paris, France	11	B1	Valve	QpH1
CB41	chronic	Human	Marseille, France	1	A2	Valve	QpDV
CB42	Chronic	Human	Toulouse, France	12	B3	Valve prosthesis	QpH1
CB43	Chronic	Human	Paris, France	7	A1	Valve	QpRS
CB44	Chronic	Human	Créteil, France	8	A2	Valve	QpRS
CB45	Chronic	Human	Paris, France	6	A2	Valve	QpRS
CB48	Abortion	Human	Grenoble, France	20	B1	Placenta	QpH1
CB50	Chronic	Human	Paris, France	20	B1	Valve prosthesis	QpH1
CB51	Abortion	Human	Madrid, Spain	4	B1	Placenta	QpDV
CB52	Chronic	Human	Paris, France	12	B3	Valve prosthesis	QpH1
CB54	Acute	Human	Aix en Provence, France	4	A2	Blood	QpDV
CB56	Chronic	Human	Paris, France	12	B3	Valve	QpH1
CB61	Chronic	Human	Marseille, France	8	A1	Valve prosthesis	QpRS
CB62	Acute	Human	Martigues, France	4	A2	Blood	QpDV
CB63	Chronic	Human	Marseille, France	1	A2	Valve	QpDV
CB64	Acute	Human	Martigues, France	1	A2	Blood	QpDV
CB65	Chronic	Human	Marseille, France	10	B1	Valve	QpRS
CB68	-	Pigeon	Marseille, France	9	B3	Stool	QpRS

CB69	Chronic	Human	Toulouse, France	13	B2	Vegetation	QpH1
CB70	Chronic	Human	Grenoble, France	8	A1	Valve	QRS
CB71	Chronic	Human	Saint Laurent du Var, France	8	A1	Valve prosthesis	QpRS
CB72	Chronic	Human	Paris, France	20	B2	Valve prosthesis	QpH1
CB74 *	Chronic	Human	Toulouse, France	14	B3	Valve prosthesis	QpH1
CB76	Chronic	Human	Broussais, France	12	A3	Valve	QpH1
CB87	Abortion	Human	Martigues, France	1	A3	Placenta	QpDV
CB92	Chronic	Human	Marseille, France	9	A3	Valve	QpRS
CB93	Abortion	Human	Dreux, France	8	B4	Placenta	QpRS
CB94	Acute	Human	Aix en Provence, France	1	B5	Blood	QpDV
CB96	Chronic	Human	Marseille, France	8	A1	Valve	QpRS
CB97	Acute	Human	Weiler, France	1	A3	blood	QpDV
CB102	Chronic	Human	Poitiers, France	15	B5	Valve prosthesis	QpH1
CB109	Chronic	Human	Germany	12	B5	Valve	QpH1
CB113	Abortion	Goat	Albi, France	12	B5	Placenta	QpH1
CB114	Chronic	Human	Marseille, France	8	A3	Valve	QpRS
CB120	Chronic	Human	Villeneuve St Georges, France	ND	B6	Blood	ND
CB125	Chronic	Human	Tours, France	ND	B5	Valve	ND
CB132	Chronic	Human	Bruxelles, Belgium	ND	A3	Blood / valve	ND
CB148	Acute - Hepatitis	Human	Marseille, France	4	A2	Blood	ND

CB151	Chronic	Human	Marseille, France	ND	A1	Valve	ND
CB154	Chronic	Human	Bruxelles, Belgium	32	B3	Lung Valve	ND
CB155	-	Soft Tick	Senegal	6	A2	Ornithodoros sonrai	ND
Derma	-	Hard Tick	Slovakia	23	B1	*Dermacentor marginatus*	QpH1
HzR	ND	Human	Italy	29	B4	Blood	QpH1
HzS	Acute	Human	Italy	18	B4	Blood	QpH1
Luga	-	Hard Tick	Leningrad, Russia	23	B1	Ixodes ricinus	QpH1
NM	-	Hard Tick	USA	16	B1	*Demacentor andersoni*	QpH1
PCat	-	Cat	Canada	21	B1	Placenta	ND
Pq	Abortion	Goat	ND	ND	B4	Placenta	ND
S217	Acute - Hepatitis	Human	USA	21	C	Liver	Less
Z2775	Abortion	Cow	Germany	16	B1	Placenta	QpH1
Z3055 *	Abortion	Goat	Germany	14	B2	Placenta	QpH1

Table 1 : Collection of isolates

* Epidemic genotype isolates

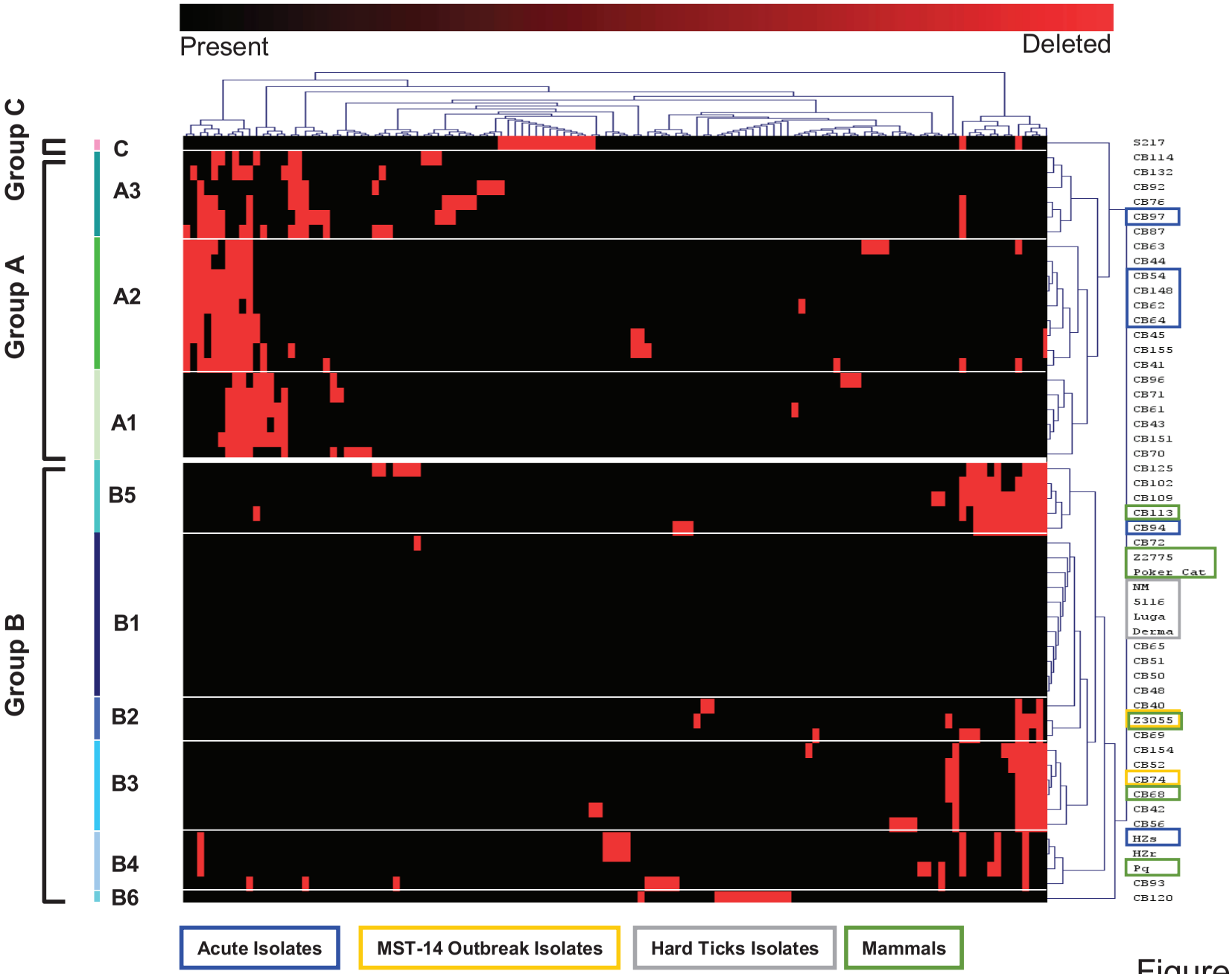
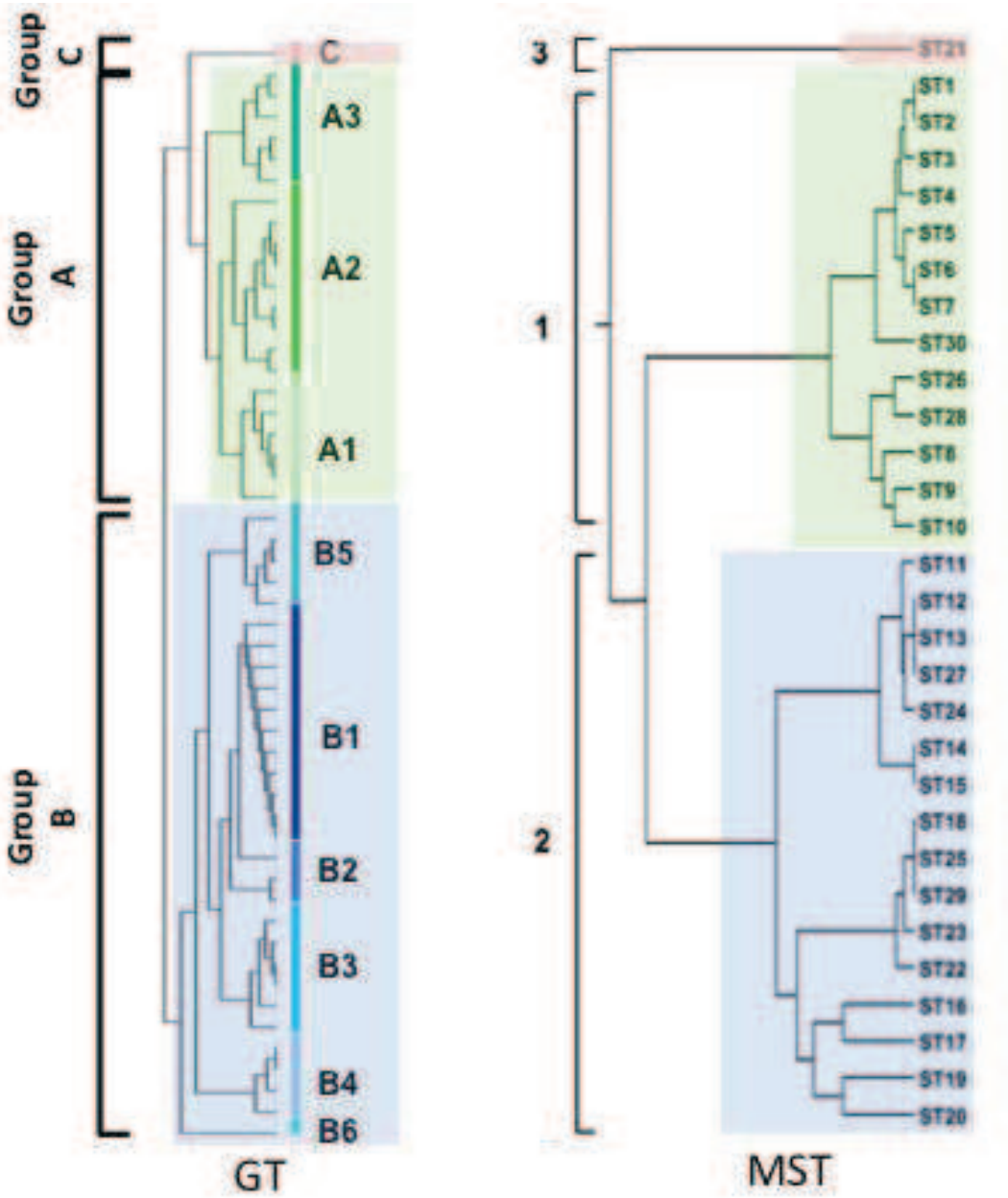

Present
Deleted
Group C
Group A
Group B
C
A3
A2
A1
B5
B1
B2
B3
B4
B6
S217
CB114
CB132
CB92
CB76
CB97
CB87
CB63
CB44
CB54
CB148
CB62
CB64
CB45
CB155
CB41
CB96
CB71
CB61
CB43
CB151
CB70
CB125
CB102
CB109
CB113
CB94
CB72
Z2775
Poker Cat
NM
5116
Luga
Derma
CB65
CB51
CB50
CB48
CB40
Z3055
CB69
CB154
CB52
CB74
CB68
CB42
CB56
HZs
HZr
Pq
CB93
CB120
Acute Isolates
MST-14 Outbreak Isolates
Hard Ticks Isolates
Mammals

Figure 1A



GT : Whole genome Typing

MST : Multi Spacer Typing

Figure 1B

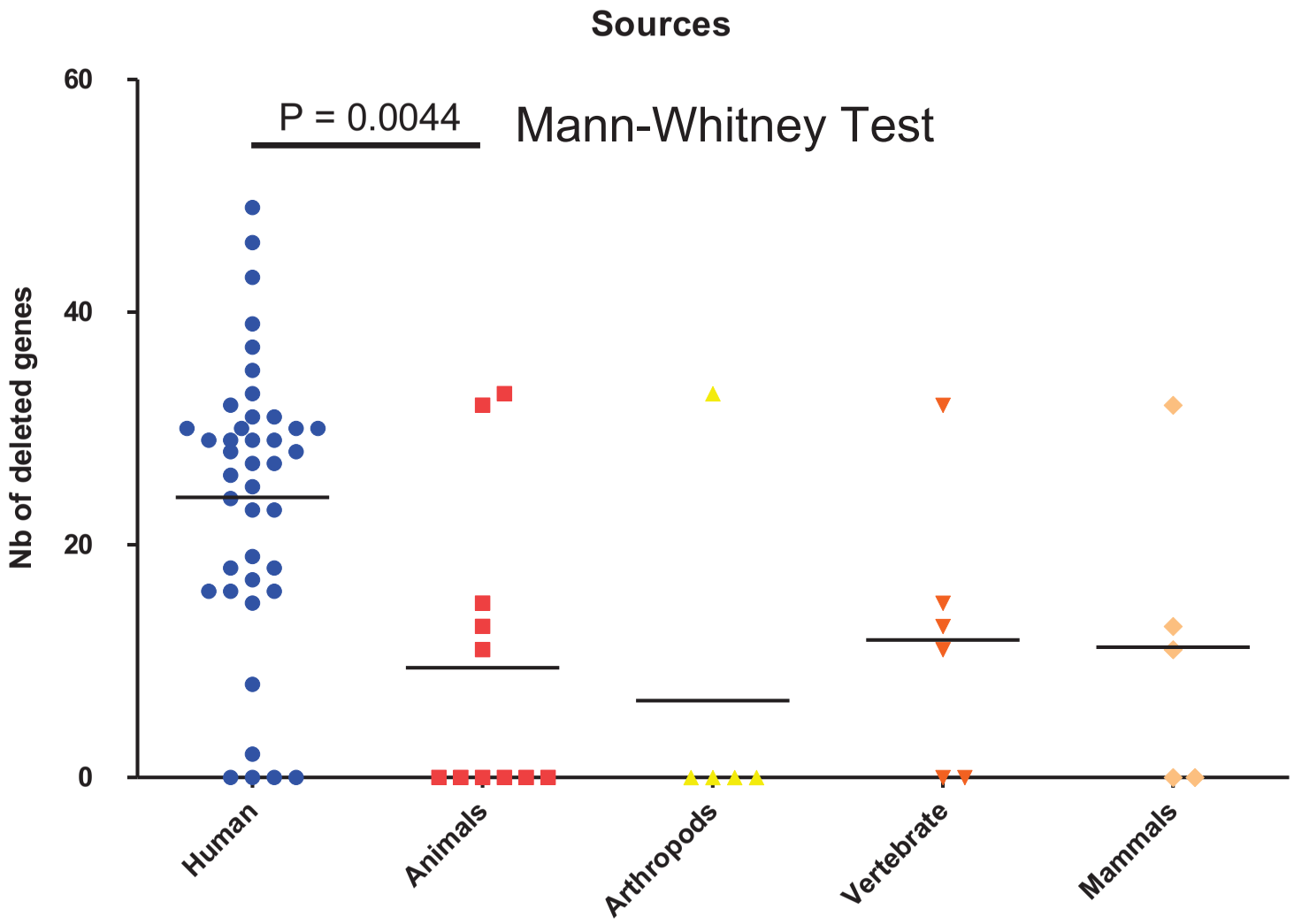

Sources
P = 0.0044
Mann-Whitney Test
Nb of deleted genes
60
40
20
0
Human
Animals
Arthropods
Vertebrate
Mammals

Figure 2A

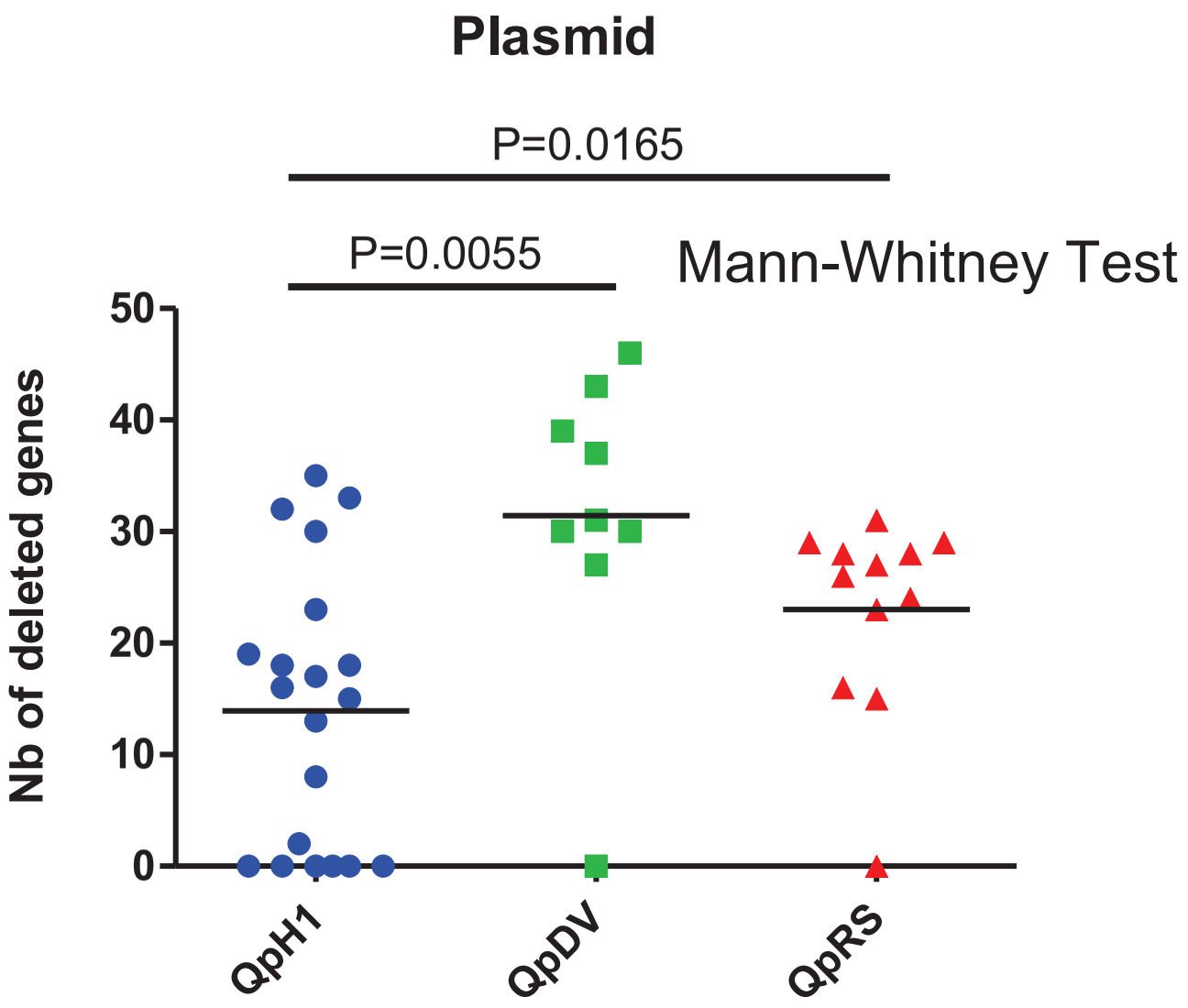

Plasmid
P=0.0165
P=0.0055
Mann-Whitney Test
Nb of deleted genes
50
40
30
20
10
0
QpH1
QpDV
QpRS

Figure 2B

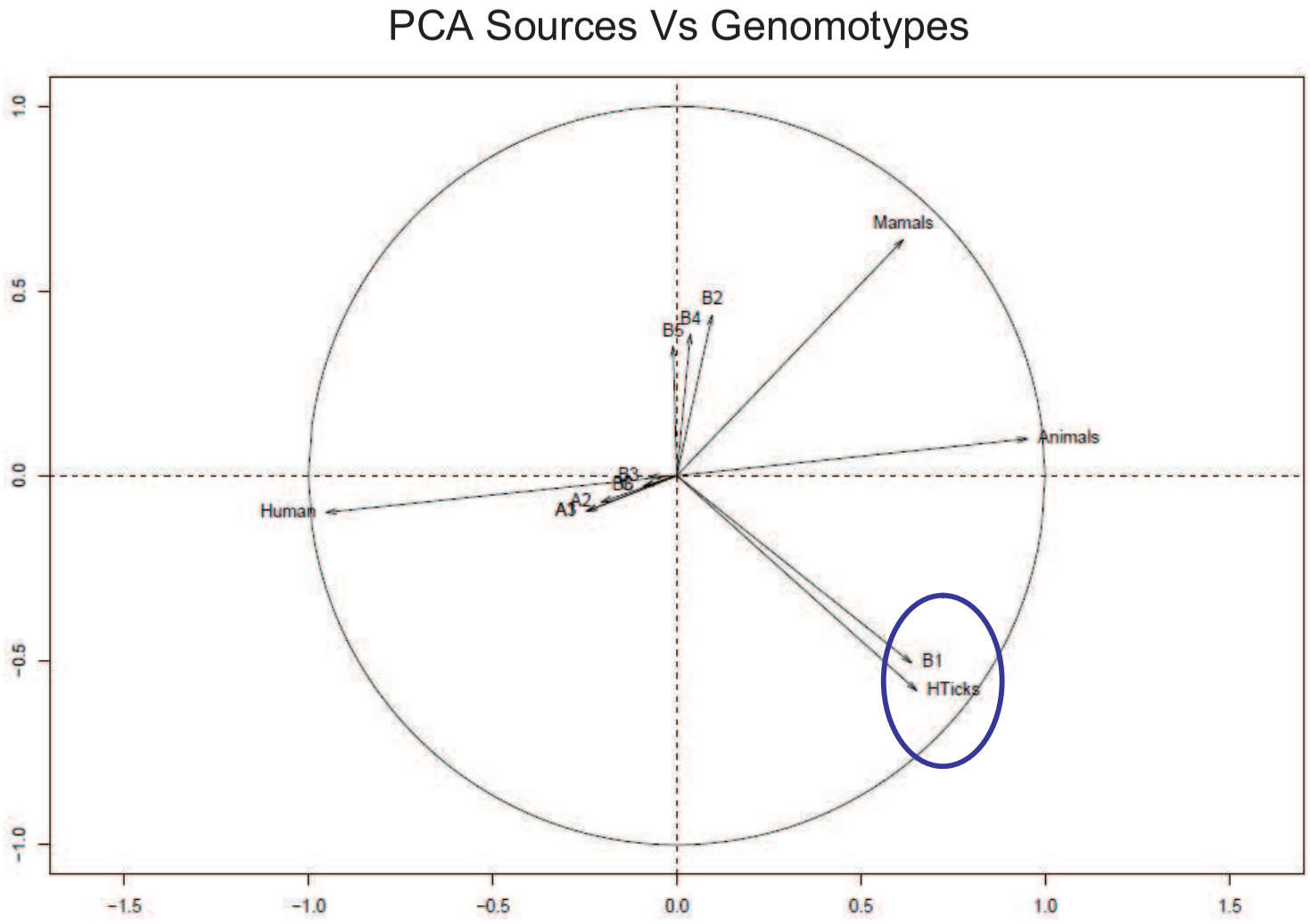
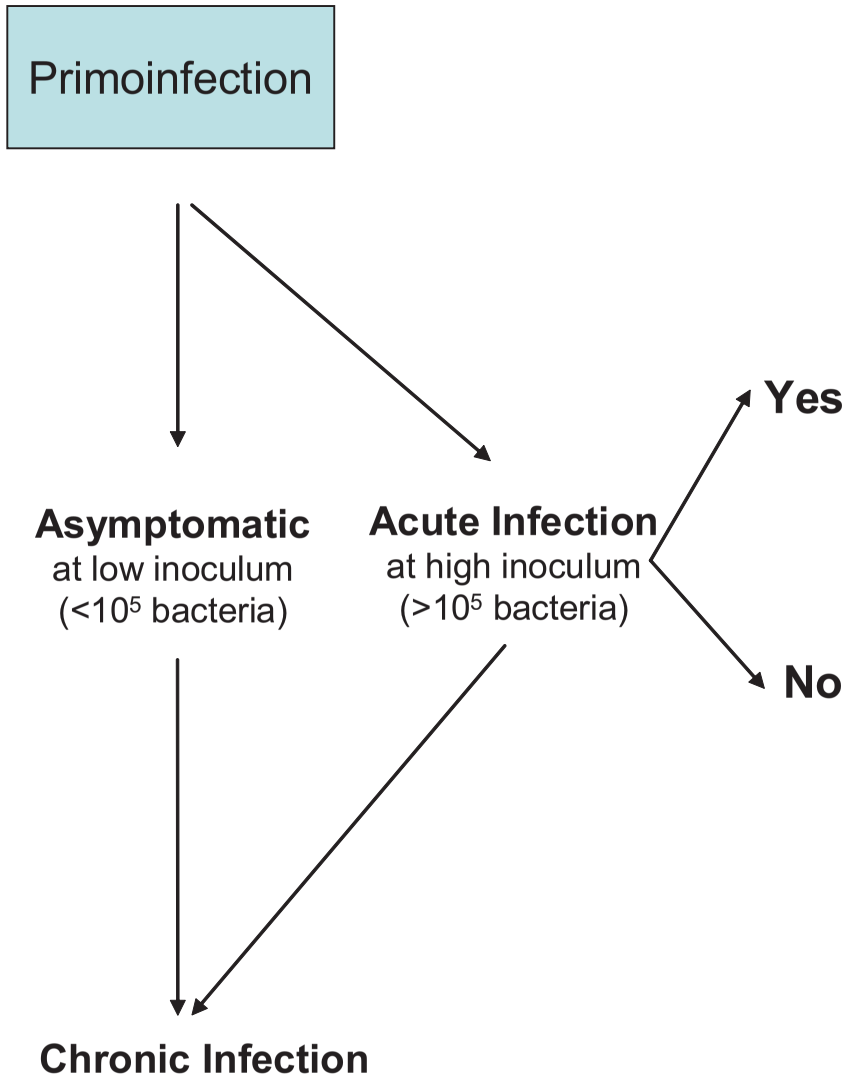

PCA Sources Vs Genomotypes
Mamals
B2
B4
B5
Animals
B3
B6
A2
A1
Human
B1
HTicks
1.0
0.5
0.0
−0.5
−1.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5

Figure 2C



Plasmid Type	MST Type	WGT Type
QpDV, QpH1	1, 2, 4, 16, 18	A2, A3, B1, B4, B5
Other plasmids and plasmid less	Other MST types	Other WGT types

Figure 3

Conclusions générales et perspectives

Cette thèse présentait deux objectifs majeurs. Le premier objectif était d'améliorer les stratégies de purification de l'ARN qui avaient été précédemment mis au point au sein du laboratoire, dans le but de permettre une analyse transcriptionnelle globale de l'hôte et de son pathogène par puce à ADN. Le second objectif était d'utiliser les puces à ADN pour entreprendre des études de transcriptomique et génomique comparative, afin d'augmenter nos connaissances sur la pathogénicité de *C. burnetii*.

Dans le champ d'étude des interactions hôte-pathogène, les analyses des profils transcriptionnels des bactéries restent très limitées comparées à celles de leur hôte du fait de limitations techniques [24]. L'URMITE est un acteur prépondérant dans la mise au point de stratégies permettant l'étude des bactéries intracellulaires strictes [65]. Pour faciliter les analyses transcriptionnelles de bactéries intracellulaires strictes, La MV. *et al*. ont mis au point un système d'hybridation soustractive permettant l'élimination des ARN eucaryotes à partir d'un modèle *in vitro* de cellules infectées par *Rickettsia conorii* [66]. Les modèles *in vitro* ne proposant qu'une vue altérée des réelles interactions entre l'hôte et son pathogène durant une infection, il était nécessaire de développer un modèle *in vivo* basé sur l'utilisation d'échantillons cliniques. Dans cette thèse, nous avons donc entrepris la mise au point d'un protocole permettant l'utilisation d'échantillons cliniques pour l'étude transcriptionnelle simultanée de l'hôte et de son pathogène, en se basant sur une stratégie similaire de séparation des ARN eucaryotes et bactériens. Contrairement à la stratégie originelle, la fraction d'ARN eucaryotes a été également conservée. Par ailleurs, la quantité des échantillons cliniques étant souvent limitée, il nous a paru nécessaire d'adapter aussi les protocoles d'extraction pour obtenir de l'ARN de qualité en quantité suffisante en vue de l'utilisation sur des puces à ADN. Pour

ce faire, nous avons mis au point notre stratégie sur un modèle de peau de lapin infectée par *R. conorii*, puis nous l'avons validé à partir d'escarres provenant de patients souffrant de la fièvre boutonneuse méditerranéenne. Au vu des premières hybridations, qui ont montré des résultats préliminaires encourageants, cette stratégie a été utilisée pour la mise en œuvre d'un projet d'étude global de la réponse transcriptionnelle de l'hôte et de son pathogène dans le cas de la fièvre boutonneuse méditerranéenne. Les fractions d'ARN eucaryotes et bactériens ont été séparées avec succès à partir d'une collection d'escarres collectées à Marseille. Les ARN de *R. conorii* ont permis d'analyser la réponse *in vivo* de la bactérie dans les escarres de patients souffrant de fièvre boutonneuse méditerranéenne [67]. Les ARN eucaryotes ont également été utilisés pour analyser la réponse de l'hôte lors d'une infection par *R. conorii*. Nous sommes actuellement en attente d'escarres complémentaires pour confirmer et valider les premiers résultats obtenus. Notre stratégie semble également être portable sur d'autres modèles *in vivo* utilisant des échantillons cliniques. Il serait à présent intéressant d'utiliser notre stratégie sur des échantillons de valves infectées par *C. burnetii* provenant de patients de fièvre Q présentant une endocardite infectieuse [32].

Comme nous l'avons évoqué précédemment, l'utilisation des puces à ADN dans le contexte des interactions hôte-pathogène reste difficile à mettre en œuvre et spécialement pour analyser les profils d'expression des procaryotes. Pour ce travail nous nous sommes appuyés sur l'expertise et l'expérience acquises lors de précédents travaux concernant des bactéries intracellulaires strictes comme *Tropheryma whipplei* [42] [68] et *R. conorii* [66,67]. Nous avons montré, lors de l'analyse transcriptomique de la réponse précoce de *C. burnetii* à un changement de température, que les profils transcriptionnels obtenus présentaient de faibles différences d'expression (Fold change max = -3,92). De

plus, nous avons pu observer que les profils transcriptionnels étaient similaires après une exposition à la chaleur ou au froid, et cela après 30 minutes ou 1 heure. Même si les variations observées étaient relativement faibles, il apparaît clairement que *C. burnetii* présente une réponse adaptée aux changements de température. Les faits les plus marquants pourraient être une accumulation de (p)ppGpp qui est un senseur du stress, un arrêt de la croissance, une modification de la membrane et de la paroi, ainsi qu'une action des ABC transporteur et des pompes à efflux. Tous ces phénomènes pourraient être associés à un passage de la bactérie en forme Small Cell Variant voire pseudo-sporulée, pour lui conférer des propriétés de résistance importante. De plus, au cours de ce travail, nous avons observé, par hasard, que les gènes différentiellement exprimés étaient pour beaucoup d'entre eux regroupés spatialement. Nos analyses bio-informatiques ont montré que ces clusters spatiaux de régulation ne répondaient ni au dogme promoteur - facteur de transcription - opéron ni à des réseaux biologiques. Nous avons retrouvé ce phénomène, passé inaperçu dans plusieurs autres études transcriptionnelles de bactéries intracellulaires [42–45,47,68], confirmant l'existence de régulations spatiales. N'ayant pu identifier les causes de régulations de ces clusters, nous posons l'hypothèse qu'elles pourraient être dues à un mode de régulation épigénétique qui reste à caractériser.

Concernant l'analyse de génomique comparative des différents isolats de *C. burnetii,* les contenus génomiques des différents isolats apparaissaient comme ultra conservés dans notre collection. Nous avons trouvé des résultats similaires à ceux précédemment trouvés par Beare et *al.* sur une collection de 24 souches de référence [56]. Des gènes divergents avaient des fréquences de délétion importantes (jusqu'à 50%), mettant en exergue des zones « hot spot » pour la plasticité du génome. Certaines de ces délétions comportaient une forte

proportion de pseudo-gènes, confirmant l'évolution réductive de ces génomes. A partir des contenus géniques de chaque isolat, nous avons effectué un génomotypage basé sur la présence ou l'absence des gènes. Ce dernier a révélée la présence de 3 groupes majeurs qui avaient déjà été identifiés par les différentes stratégies de typage et spécialement avec le typage des séquences intergéniques (MST) [55]. Cela témoigne d'une co-évolution entre les souches de ces différents groupes. L'analyse de deux isolats, présentant un génotype identique majoritaire lors de l'épidémie qui a sévi aux Pays-Bas [40], ne nous pas permis de conclure à l'existence d'un clone épidémique. Enfin, l'analyse de l'association des informations physiopathologiques et du contenu génique nous a permis d'améliorer nos connaissances sur les facteurs bactériens associés la fièvre Q aiguë. De plus, nous avons trouvé que les isolats humains contenaient en moyenne plus d'évènements de délétion que les isolats d'origine animale qui eux étaient presque totalement associés au plasmide QpH1. Ce phénomène a particulièrement été observé chez les isolats provenant de tiques dures. Ces derniers ne présentent aucune différence avec la souche de référence Nine Mile I. Nous avons également remarqué que certains isolats d'origine animale et humaine présentaient également un contenu génique identique à celui des tiques dures. Bien que la fièvre Q fût initialement décrite comme une maladie vectorielle [33,34], la voie principale de dissémination considérée reste à ce jour la contamination par aérosols [69]. Au vu de nos résultats, et considérant aussi la prévalence importante de *C. burnetii* dans les arthropodes [35], la transmission vectorielle de la fièvre Q est certainement sous-estimée. Les tiques dures pourraient également être impliquées dans la distribution de clones spécifiques.

Nous avons montré au cours de cette thèse que l'utilisation des puces à ADN permettait d'augmenter nos connaissances sur la pathogénicité de l'agent de la fièvre Q. Cependant, nos travaux restent soumis à caution. En effet, concernant l'analyse du transcriptome, le génome de *C. burnetii* étant encore

peu analysé [11,37] (plus de la moitié des gènes sont annotés « unknown protein »), nous n'avons pu que partiellement interpréter la réponse de *C. burnetii* aux changements de température. De plus, nous avons utilisé une puce à ADN à faible densité (1 sonde par gène déposée). Ce manque de résolution nous a empêchés de visualiser l'existence d'ARNnc et leur potentielle régulation. Concernant l'étude de génomique comparative, les puces à ADN reposant sur le principe d'hybridation, toutes séquences inexistantes dans le génome de Nine Mile n'ont pu être identifiées et intégrées aux résultats. Nous nous sommes donc limités aux évènements de délétion chromosomique. Enfin, comme pour l'analyse du transcriptome, la puce à faible résolution utilisée ne permet pas d'avoir une forte précision dans les résultats. L'utilisation d'outils plus résolutifs pourrait permettre de confirmer et d'affiner nos résultats.

L'amélioration des technologies de séquençage, avec l'apparition des séquenceurs à haut débit, fait du séquençage une alternative crédible à l'utilisation des puces à ADN. Même si son coût reste nettement supérieur, le séquençage de l'ADNg et l'ARN par séquenceur à haut débit (DNAseq et RNAseq) comporte de gros avantages comparés à l'utilisation des puces à ADN. Concernant l'utilisation des puces à ADN pour la transcriptomique, les niveaux d'expression observés sont relatifs à la quantité de fluorescence observée. Il est à noter qu'il est impossible de comparer ces quantités relatives entre les différentes sondes du fait de propriétés thermodynamiques différentes [21,70]. Le RNAseq présente l'avantage d'être *a priori* un outil quantitatif qui pourra permettre de connaître les réels niveaux d'expression des gènes. Enfin, même si l'appariation des tiling-microarrays (puces à ADN comprenant l'ensemble des nucléotides d'un génome) permet d'obtenir une résolution élevée, elle ne résout pas le problème des séquences non spottées sur la lame, contrairement au séquençage.

Pourtant, aussi prometteuses que soient le DNAseq et RNAseq, ils comporteront les mêmes limitations que pour les puces à ADN dans un système de bactéries intracellulaires. Les ADN ou ARN contaminants doivent être éliminés dans le but d'éviter qu'ils soient séquencés et qu'ils diminuent la profondeur du séquençage. Il est donc nécessaire de développer de nouvelles stratégies pour séparer les ARN eucaryotes et procaryotes. Les stratégies déjà développées pour les puces à ADN semblent pouvoir être adaptées aux technologies DNAseq et RNAseq.

A ce jour, quelques études de génomique comparative et transcriptomique utilisant le séquençage à haut débit ont été réalisées pour des bactéries intracellulaires strictes que sont *C. burnetii* [37] et *Chlamydia trachomatis* [71]. Ces études ont pu mettre en évidence des variations et remaniements chromosomiques ainsi que du polymorphisme nucléotidique entre 4 souches de *C. burnetii* pour le DNAseq et des régulations de l'expression génétique, l'identification d'ARNnc, d'identifier les structures primaires des ARN de *Chlamydia trachomatis* pour le RNAseq. L'utilisation de ces deux nouvelles technologies pourrait donc rapidement accroitre les connaissances concernant la pathogénicité des bactéries intracellulaires strictes et remplacer à terme l'utilisation des puces à ADN. Dans le laboratoire, nous disposons actuellement de deux séquenceurs haut-débit (Roche 454 et Applied Solid). Au sein du laboratoire, le DNAseq est déjà largement utilisé pour le séquençage *de novo* de nouvelles espèces bactériennes et le RNAseq est en cours de développement.

References

1. Tamas I, Klasson LM, Sandstrom JP, Andersson SG (2001) Mutualists and parasites: how to paint yourself into a (metabolic) corner. FEBS Lett 498: 135-139. S0014-5793(01)02459-0 [pii].

2. Raoult D, Roux V (1997) Rickettsioses as paradigms of new or emerging infectious diseases. Clin Microbiol Rev 10: 694-719.

3. Storz J, Spears P (1977) Chlamydiales: properties, cycle of development and effect on eukaryotic host cells. Curr Top Microbiol Immunol 76: 167-214.

4. Paddock CD, Childs JE (2003) *Ehrlichia chaffeensis*: a prototypical emerging pathogen. Clin Microbiol Rev 16: 37-64.

5. Dumler JS (2005) Anaplasma and Ehrlichia infection. Ann N Y Acad Sci 1063: 361-373. 1063/1/361 [pii];10.1196/annals.1355.069 [doi].

6. Walker SL, Lockwood DN (2007) Leprosy. Clin Dermatol 25: 165-172. S0738-081X(06)00066-6 [pii];10.1016/j.clindermatol.2006.05.012 [doi].

7. Maurin M, Raoult D (1999) Q fever. Clin Microbiol Rev 12: 518-553.

8. Hackstadt T (1998) The diverse habitats of obligate intracellular parasites. Curr Opin Microbiol 1: 82-87. S1369-5274(98)80146-X [pii].

9. Andersson SG, Zomorodipour A, Andersson JO, Sicheritz-Ponten T, Alsmark UC, Podowski RM, Naslund AK, Eriksson AS, Winkler HH, Kurland CG (1998) The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. Nature 396: 133-140. 10.1038/24094 [doi].

10. Ogata H, Audic S, Renesto-Audiffren P, Fournier PE, Barbe V, Samson D, Roux V, Cossart P, Weissenbach J, Claverie JM, Raoult D (2001) Mechanisms of evolution in *Rickettsia conorii* and *R. prowazekii*. Science 293: 2093-2098. 10.1126/science.1061471 [doi];293/5537/2093 [pii].

11. Seshadri R, Paulsen IT, Eisen JA, Read TD, Nelson KE, Nelson WC, Ward NL, Tettelin H, Davidsen TM, Beanan MJ, Deboy RT, Daugherty SC, Brinkac LM, Madupu R, Dodson RJ, Khouri HM, Lee KH, Carty HA, Scanlan D, Heinzen RA, Thompson HA, Samuel JE, Fraser CM, Heidelberg JF (2003) Complete genome sequence of the Q-fever pathogen *Coxiella burnetii*. Proc Natl Acad Sci U S A 100: 5455-5460. 10.1073/pnas.0931379100 [doi];0931379100 [pii].

12. Colston MJ (2001) Mycobacterium leprae genome sequence; a landmark achievement. Lepr Rev 72: 385-386.

13. Saenz HL, Dehio C (2005) Signature-tagged mutagenesis: technical advances in a negative selection method for virulence gene identification. Curr Opin Microbiol 8: 612-619. S1369-5274(05)00129-3 [pii];10.1016/j.mib.2005.08.013 [doi].

14. Wood DO, Azad AF (2000) Genetic manipulation of rickettsiae: a preview. Infect Immun 68: 6091-6093.

15. Binet R, Maurelli AT (2009) Transformation and isolation of allelic exchange mutants of Chlamydia psittaci using recombinant DNA introduced by electroporation. Proc Natl Acad Sci U S A 106: 292-297. 0806768106 [pii];10.1073/pnas.0806768106 [doi].

16. Driskell LO, Yu XJ, Zhang L, Liu Y, Popov VL, Walker DH, Tucker AM, Wood DO (2009) Directed mutagenesis of the Rickettsia prowazekii pld gene encoding phospholipase D. Infect Immun 77: 3244-3248. IAI.00395-09 [pii];10.1128/IAI.00395-09 [doi].

17. Beare PA, Howe D, Cockrell DC, Omsland A, Hansen B, Heinzen RA (2009) Characterization of a Coxiella burnetii ftsZ mutant generated by Himar1 transposon mutagenesis. J Bacteriol 191: 1369-1381. JB.01580-08 [pii];10.1128/JB.01580-08 [doi].

18. Fournier PE, Drancourt M, Raoult D (2007) Bacterial genome sequencing and its use in infectious diseases. Lancet Infect Dis 7: 711-723. S1473-3099(07)70260-8 [pii];10.1016/S1473-3099(07)70260-8 [doi].

19. Gaywee J, Radulovic S, Higgins JA, Azad AF (2002) Transcriptional analysis of *Rickettsia prowazekii* invasion gene homolog (invA) during host cell infection. Infect Immun 70: 6346-6354.

20. Rovery C, La MV, Robineau S, Matsumoto K, Renesto P, Raoult D (2005) Preliminary transcriptional analysis of spoT gene family and of membrane proteins in *Rickettsia conorii* and *Rickettsia felis*. Ann N Y Acad Sci 1063: 79-82. 1063/1/79 [pii];10.1196/annals.1355.011 [doi].

21. Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270: 467-470.

22. Ehrenreich A (2006) DNA microarray technology for the microbiologist: an overview. Appl Microbiol Biotechnol 73: 255-273. 10.1007/s00253-006-0584-2 [doi].

23. Cummings CA, Relman DA (2000) Using DNA microarrays to study host-microbe interactions. Emerg Infect Dis 6: 513-525.

24. La MV, Raoult D, Renesto P (2008) Regulation of whole bacterial pathogen transcription within infected hosts. FEMS Microbiol Rev 32: 440-460. FMR103 [pii];10.1111/j.1574-6976.2008.00103.x [doi].

25. Rovery C, Raoult D (2008) Mediterranean spotted fever. Infect Dis Clin North Am 22: 515-30, ix. S0891-5520(08)00022-6 [pii];10.1016/j.idc.2008.03.003 [doi].

26. Maurin M, Benoliel AM, Bongrand P, Raoult D (1992) Phagolysosomes of Coxiella burnetii-infected cell lines maintain an acidic pH during persistent infection. Infect Immun 60: 5013-5016.

27. Hackstadt T, Peacock MG, Hitchcock PJ, Cole RL (1985) Lipopolysaccharide variation in Coxiella burnetti: intrastrain heterogeneity in structure and antigenicity. Infect Immun 48: 359-365.

28. Hackstadt T (1990) The role of lipopolysaccharides in the virulence of Coxiella burnetii. Ann N Y Acad Sci 590: 27-32.

29. Hoover TA, Culp DW, Vodkin MH, Williams JC, Thompson HA (2002) Chromosomal DNA deletions explain phenotypic characteristics of two antigenic variants, phase II and RSA 514 (crazy), of the *Coxiella burnetii* Nine Mile strain. Infect Immun 70: 6726-6733.

30. Heinzen RA, Hackstadt T, Samuel JE (1999) Developmental biology of Coxiella burnettii. Trends Microbiol 7: 149-154. S0966842X99014754 [pii].

31. Coleman SA, Fischer ER, Howe D, Mead DJ, Heinzen RA (2004) Temporal analysis of Coxiella burnetii morphological differentiation. J Bacteriol 186: 7344-7352. 186/21/7344 [pii];10.1128/JB.186.21.7344-7352.2004 [doi].

32. Raoult D, Marrie T, Mege J (2005) Natural history and pathophysiology of Q fever. Lancet Infect Dis 5: 219-226. S1473-3099(05)70052-9 [pii];10.1016/S1473-3099(05)70052-9 [doi].

33. Derrick EH (1937) "Q" fever, a new fever entity: clinical features, diagnosis and laboratory investigation. Med J Australia 2: 281-299.

34. Derrick EH (1973) The course of infection with *Coxiella burnetii*. Med J Australia 1: 1051-1057.

35. Mediannikov O, Fenollar F, Socolovschi C, Diatta G, Bassene H, Molez JF, Sokhna C, Trape JF, Raoult D (2010) Coxiella burnetii in humans and ticks in rural Senegal. PLoS Negl Trop Dis 4: e654. 10.1371/journal.pntd.0000654 [doi].

36. Enserink M (2010) Infectious diseases. Questions abound in Q-fever explosion in the Netherlands. Science 327: 266-267. 327/5963/266-a [pii];10.1126/science.327.5963.266-a [doi].

37. Beare PA, Unsworth N, Andoh M, Voth DE, Omsland A, Gilk SD, Williams KP, Sobral BW, Kupko JJ, III, Porcella SF, Samuel JE, Heinzen RA (2009) Comparative genomics reveal extensive transposon-mediated genomic plasticity and diversity among potential effector proteins within the genus Coxiella. Infect Immun 77: 642-656. IAI.01141-08 [pii];10.1128/IAI.01141-08 [doi].

38. Voth DE, Howe D, Beare PA, Vogel JP, Unsworth N, Samuel JE, Heinzen RA (2009) The *Coxiella burnetii* ankyrin repeat domain-containing protein family is heterogeneous, with C-terminal truncations that influence Dot/Icm-mediated secretion. J Bacteriol 191: 4232-4242. JB.01656-08 [pii];10.1128/JB.01656-08 [doi].

39. Raoult D (2009) Reemergence of Q fever after 11 September 2001. Clin Infect Dis 48: 558-559. 10.1086/596706 [doi];10.1086/596706 [pii].

40. Klaassen CH, Nabuurs-Franssen MH, Tilburg JJ, Hamans MA, Horrevorts AM (2009) Multigenotype Q fever outbreak, the Netherlands. Emerg Infect Dis 15: 613-614.

41. Klinkert B, Narberhaus F (2009) Microbial thermosensors. Cell Mol Life Sci 66: 2661-2676. 10.1007/s00018-009-0041-3 [doi].

42. Crapoulet N, Barbry P, Raoult D, Renesto P (2006) Global transcriptome analysis of Tropheryma whipplei in response to temperature stresses. J Bacteriol 188: 5228-5239. 188/14/5228 [pii];10.1128/JB.00507-06 [doi].

43. Dreher-Lesnick SM, Ceraul SM, Rahman MS, Azad AF (2008) Genome-wide screen for temperature-regulated genes of the obligate intracellular bacterium, *Rickettsia typhi*. BMC Microbiol 8: 61. 1471-2180-8-61 [pii];10.1186/1471-2180-8-61 [doi].

44. Ellison DW, Clark TR, Sturdevant DE, Virtaneva K, Hackstadt T (2009) Limited transcriptional responses of *Rickettsia rickettsii* exposed to environmental stimuli. PLoS One 4: e5612. 10.1371/journal.pone.0005612 [doi].

45. Audia JP, Patton MC, Winkler HH (2008) DNA microarray analysis of the heat shock transcriptome of the obligate intracytoplasmic pathogen *Rickettsia prowazekii*. Appl Environ Microbiol 74: 7809-7812. AEM.00896-08 [pii];10.1128/AEM.00896-08 [doi].

46. Toledo-Arana A, Repoila F, Cossart P (2007) Small noncoding RNAs controlling pathogenesis. Curr Opin Microbiol 10: 182-188. S1369-5274(07)00021-5 [pii];10.1016/j.mib.2007.03.004 [doi].

47. Toledo-Arana A, Dussurget O, Nikitas G, Sesto N, Guet-Revillet H, Balestrino D, Loh E, Gripenland J, Tiensuu T, Vaitkevicius K, Barthelemy M, Vergassola M, Nahori MA, Soubigou G, Regnault B, Coppee JY, Lecuit M, Johansson J, Cossart P (2009) The Listeria transcriptional landscape from saprophytism to virulence. Nature 459: 950-956. nature08080 [pii];10.1038/nature08080 [doi].

48. Faucher SP, Friedlander G, Livny J, Margalit H, Shuman HA (2010) Legionella pneumophila 6S RNA optimizes intracellular multiplication. Proc Natl Acad Sci U S A 107: 7533-7538. 0911764107 [pii];10.1073/pnas.0911764107 [doi].

49. Passalacqua KD, Varadarajan A, Ondov BD, Okou DT, Zwick ME, Bergman NH (2009) Structure and complexity of a bacterial transcriptome. J Bacteriol 191: 3203-3211. JB.00122-09 [pii];10.1128/JB.00122-09 [doi].

50. Stein A, Raoult D (1993) Lack of pathotype specific gene in human Coxiella burnetii isolates. Microb Pathog 15: 177-185. S0882-4010(83)71068-5 [pii];10.1006/mpat.1993.1068 [doi].

51. Hendrix LR, Samuel JE, Mallavia LP (1991) Differentiation of Coxiella burnetii isolates by analysis of restriction-endonuclease-digested DNA separated by SDS-PAGE. J Gen Microbiol 137: 269-276.

52. Jager C, Willems H, Thiele D, Baljer G (1998) Molecular characterization of Coxiella burnetii isolates. Epidemiol Infect 120: 157-164.

53. Nguyen SV, Hirai K (1999) Differentiation of Coxiella burnetii isolates by sequence determination and PCR-restriction fragment length polymorphism analysis of isocitrate dehydrogenase gene. FEMS Microbiol Lett 180: 249-254. S0378-1097(99)00495-4 [pii].

54. Andoh M, Nagaoka H, Yamaguchi T, Fukushi H, Hirai K (2004) Comparison of Japanese isolates of Coxiella burnetii by PCR-RFLP and sequence analysis. Microbiol Immunol 48: 971-975. JST.JSTAGE/mandi/48.971 [pii].

55. Glazunova O, Roux V, Freylikman O, Sekeyova Z, Fournous G, Tyczka J, Tokarevich N, Kovacava E, Marrie TJ, Raoult D (2005) Coxiella burnetii genotyping. Emerg Infect Dis 11: 1211-1217.

56. Beare PA, Samuel JE, Howe D, Virtaneva K, Porcella SF, Heinzen RA (2006) Genetic diversity of the Q fever agent, *Coxiella burnetii*, assessed by microarray-based whole-genome comparisons. J Bacteriol 188: 2309-2324. 188/7/2309 [pii];10.1128/JB.188.7.2309-2324.2006 [doi].

57. Stein A, Lepidi H, Mege JL, Marrie TJ, Raoult D (2000) Repeated pregnancies in BALB/c mice infected with Coxiella burnetii cause disseminated infection, resulting in stillbirth and endocarditis. J Infect Dis 181: 188-194. JID990157 [pii];10.1086/315166 [doi].

58. Russell-Lodrigue KE, Zhang GQ, McMurray DN, Samuel JE (2006) Clinical and pathologic changes in a guinea pig aerosol challenge model of acute Q fever. Infect Immun 74: 6085-6091. 74/11/6085 [pii];10.1128/IAI.00763-06 [doi].

59. Russell-Lodrigue KE, Andoh M, Poels MW, Shive HR, Weeks BR, Zhang GQ, Tersteeg C, Masegi T, Hotta A, Yamaguchi T, Fukushi H, Hirai K, McMurray DN, Samuel JE (2009) Coxiella burnetii isolates cause genogroup-specific virulence in mouse and guinea pig models of acute Q fever. Infect Immun 77: 5640-5650. IAI.00851-09 [pii];10.1128/IAI.00851-09 [doi].

60. Weyer F (1953) Die Bezichungen des Q-Fieber Erregers (Rickettsia burneti) zu Arthropoden. Z Tropenmed Parasitol 4(3): 344-382.

61. PHILIP CB (1948) Observations on experimental Q fever. J Parasitol 34: 457-464.

62. Balashov, Y. S. and Daiter, A. B. (1973) Blood-feeding artropods and rickettsiae [In Russian]. Leningrad: Nauka (Russian Academy of Sciences). 251 p.

63. Smith DJW, Brown HE, Derrick EH (1939) Further series of laboratoriy infections with the Rickettsia of "Q" fever. Med J Australia 1: 13-20.

64. Smith DJW, Derrick EH (1940) Studies in the epidemiology of Q fever. I. The isolation of six strains of *Rickettsia burneti* from the tick, *Haemaphysalis humerosa*. Aust J Exp Biol Med Sci 18: 1.

65. Renesto P, Crapoulet N, Ogata H, La SB, Vestris G, Claverie JM, Raoult D (2003) Genome-based design of a cell-free culture medium for *Tropheryma whipplei*. Lancet 362: 447-449. S0140-6736(03)14071-8 [pii];10.1016/S0140-6736(03)14071-8 [doi].

66. La MV, Francois P, Rovery C, Robineau S, Barbry P, Schrenzel J, Raoult D, Renesto P (2007) Development of a method for recovering rickettsial RNA from infected cells to analyze gene expression profiling of obligate intracellular bacteria. J Microbiol Methods 71: 292-297. S0167-7012(07)00325-9 [pii];10.1016/j.mimet.2007.09.017 [doi].

67. Renesto P, Rovery C, Schrenzel J, Leroy Q, Huyghe A, Li W, Lepidi H, Francois P, Raoult D (2008) *Rickettsia conorii* transcriptional response within inoculation eschar. PLoS One 3: e3681. 10.1371/journal.pone.0003681 [doi].

68. Van LM, Barbry P, Raoult D, Renesto P (2007) Molecular basis of *Tropheryma whipplei* doxycycline susceptibility examined by transcriptional profiling. J Antimicrob Chemother 59: 370-377. dkl507 [pii];10.1093/jac/dkl507 [doi].

69. Derrick EH (1983) "Q: fever, a new fever entity: clinical features, diagnosis and laboratory investigation. Rev Infect Dis 5: 790-800.

70. Shalon D, Smith SJ, Brown PO (1996) A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. Genome Res 6: 639-645.

71. Albrecht M, Sharma CM, Reinhardt R, Vogel J, Rudel T (2010) Deep sequencing-based discovery of the Chlamydia trachomatis transcriptome. Nucleic Acids Res 38: 868-877. gkp1032 [pii];10.1093/nar/gkp1032 [doi].

Annexes

Au cours de cette thèse, du fait de l'expertise acquise sur les méthodes de post-génomiques, nous avons été amenés à travailler sur d'autres bactéries intracellulaires et spécialement des Rickettsies à savoir *Rickettisa prowazekii*, *Rickettsia africae* et *R. conorii*. Ces dernières sont des alpha-protéobactéries à petit génome en cours d'évolution réductive dont certaines ont été séquencées et ont permis d'entreprendre des analyses à l'aide d'outils de post-génomique et pour la majorité avec les puces ADN. Les annexes comportent certains de ces récents travaux.

L'annexe 1 est une analyse transcriptionnelle de *R. conorii* à partir d'escarres obtenues sur des patients atteints de fièvre boutonneuse méditerranéenne. L'analyse transcriptionnelle met en évidence un profil conservé entre les différentes escarres mais également un arrêt de la croissance bactérienne. Cette étude est à ce jour la seule étude transcriptomique par puce à ADN utilisant un modèle *in vivo* à partir d'échantillons cliniques sur les Rickettsies.

L'annexe 2 est le séquençage et l'analyse du génome de *R. africae* agent responsable de la fièvre boutonneuse d'Afrique. Ce travail a notamment montré entre autre que le génome de *R. africae* comportait un grand nombre de gènes *spoT* responsables de la synthèse et de la dégradation de l'alarmone ((p)ppGpp). Une analyse transcriptionnelle de ces gènes a pu montrer que ces derniers étaient fonctionnels et régulés lors de changements de température.

L'annexe 3 est une analyse multi-omique comparative (génomique, transcriptomique et protéomique) de deux souches de *R. prowazekii*, l'agent responsable du typhus épidémique. Ces deux souches présentent un phénotype de virulence différent à savoir une souche vaccinale avirulente et une souche pathogène isolée d'un patient. L'analyse transcriptionnelle de ces deux souches dans un modèle *in vitro* a montré notamment des variations transcriptionnelles.

Annexe 1

Rickettsia conorii Transcriptional Response within Inoculation Eschar

Patricia RENESTO[1], Clarisse ROVERY[1], Jacques SCHRENZEL[2], Quentin LEROY[1], Antoine HUYGHE[2], Wenjun LI[1], Hubert LEPIDI[1], Patrice FRANCOIS[2], Didier RAOULT[1]

Plos One

2008;3(11):e3681. Epub 2008 Nov 10

1 - Unité des Rickettsies, IRD-CNRS UMR 6236, Faculté de Médecine, Marseille, France

2 - Service of Infectious Diseases/Genomic Research Laboratory, University of GenevaHospitals, Geneva, Switzerland

OPEN ACCESS Freely available online

PLoS one

Rickettsia conorii Transcriptional Response within Inoculation Eschar

Patricia Renesto[1]*, Clarisse Rovery[1], Jacques Schrenzel[2], Quentin Leroy[1], Antoine Huyghe[2], Wenjun Li[1], Hubert Lepidi[1], Patrice François[2], Didier Raoult[1]

1 Unité des Rickettsies, IRD-CNRS UMR 6236, Faculté de Médecine, Marseille, France, **2** Service of Infectious Diseases/Genomic Research Laboratory, University of Geneva Hospitals, Geneva, Switzerland

Abstract

Background: *Rickettsia conorii*, the causative agent of the Mediterranean spotted fever, is transmitted to humans by the bite of infected ticks *Rhipicephalus sanguineus*. The skin thus constitutes an important barrier for the entry and propagation of *R. conorii*. Given this, analysis of the survival strategies used by the bacterium within infected skin is critical for our understanding of rickettsiosis.

Methodology/Principal Findings: Here, we report the first genome-wide analysis of *R. conorii* gene expression from infected human skin biopsies. Our data showed that *R. conorii* exhibited a striking transcript signature that is remarkably conserved across patients, regardless of genotype. The expression profiles obtained using custom Agilent microarrays were validated by quantitative RT-PCR. Within eschars, the amount of detected *R. conorii* transcripts was of 55%, this value being of 74% for bacteria grown in Vero cells. In such infected host tissues, approximately 15% ($n = 211$) of the total predicted *R. conorii* ORFs appeared differentially expressed compared to bacteria grown in standard laboratory conditions. These genes are mostly down-regulated and encode proteins essential for bacterial replication. Some of the strategies displayed by rickettsiae to overcome the host defense barriers, thus avoiding killing, were also pointed out. The observed up-regulation of rickettsial genes associated with DNA repair is likely to correspond to a DNA-damaging agent enriched environment generated by the host cells to eradicate the pathogens. Survival of *R. conorii* within eschars also involves adaptation to osmotic stress, changes in cell surface proteins and up-regulation of some virulence factors. Interestingly, in contrast to down-regulated transcripts, we noticed that up-regulated ones rather exhibit a small nucleotide size, most of them being exclusive for the spotted fever group rickettsiae.

Conclusion/Significance: Because eschar is a site for rickettsial introduction, the pattern of rickettsial gene expression observed here may define how rickettsiae counteract the host defense.

Citation: Renesto P, Rovery C, Schrenzel J, Leroy Q, Huyghe A, et al. (2008) *Rickettsia conorii* Transcriptional Response within Inoculation Eschar. PLoS ONE 3(11): e3681. doi:10.1371/journal.pone.0003681

Editor: Aric Gregson, University of California Los Angeles, United States of America

Received July 23, 2008; **Accepted** October 20, 2008; **Published** November 10, 2008

Funding: This project was granted by a research program "Hospital Clinical Research Program" from the French Health Ministry. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: patricia.renesto@medecine.univ-mrs.fr

Introduction

Rickettsia conorii is a Gram-negative bacterium responsible for the Mediterranean spotted fever (MSF), a disease transmitted to humans by the brown dog tick *Rhipicephalus sanguineus* [1]. Inoculation of rickettsiae to human beings leads to vasculitis and lesions at the site of tick bite [2]. The cutaneous necrosis that results from severe injury to many small vessels and otherwise called the "tâche noire" is the hallmark of many spotted fever group rickettsioses [3]. The histological examination of eschars collected from patients suffering from boutonneuse fever indicated that the alterations were mainly located in the dermis and subcutaneous tissues and evidenced the presence of rickettsiae in blood vessels [3]. Thus, indirect immunofluorescent detection of *R. conorii* on cryostat sections of skin biopsy specimens from patients was found to improve the early diagnosis of severe and atypical forms of MSF [4,5].

Inoculation eschars correspond to the portal of entry of the infectious agent into the host and the first site of challenge between the infected human being and the bacterium. Within the first 24 hours after the tick attachment, the rickettsiae are already bloodborne and the observed rickettsiae in the tick feeding site and in particular within the eschar are left over rickettsiae undergoing clearance [2]. In this respect, the "tâche noire" was depicted as being an excellent, accessible model for the study of the human-*Rickettsia* interaction [6]. The intralesional expression of local mediators of inflammation and of immune response that could contribute both to anti-rickettsial immunity and the pathogenesis of the MSF, has recently been depicted in infected human tissues [7]. Analysis of the complementary picture, namely the survival strategies used by *R. conorii* within the inoculation site should provide a better understanding of rickettsial pathogenesis. While reports of global gene expression profiling in human tissue or non-invasive patient samples suffering from bacterial diseases are limited [8,9], we explored the RNA profiles of *R. conorii* from eschars collected on MSF patients. This study was made possible by applying a strategy combining removal of eukaryotic contaminants

Table 1. Clinical characteristics of patients included in this study.

sample	age	sex	number of eschars	year of eschar sampling	severe form	geographical site of bite	MST genotype	strain isolated
E	62	F	1	2005	No	France (13)	Nd	no
G	42	F	1	2005	No	Marocco	C	no
C	72	F	1	2004	Yes	France (84)	A	yes
I	67	M	1	2005	Yes	France (13)	A	no
F,D	61	F	2	2004	Yes	France (13)	A	yes
A,B	45	M	1	2006	No	France (13)	A	yes
H	30	M	1	2005	No	Algeria	B	yes

(nd) not determined.
doi:10.1371/journal.pone.0003681.t001

with subsequent random amplification of prokaryotic cDNA [10] that was found convenient for microarray-based transcriptome analysis of obligate intracellular rickettsiae [11]. To identify genes differentially regulated within eschars, rickettsial microarrays hybridized with cDNA obtained from *R. conorii* grown in Vero cells monolayers at 32°C were used as control. The results obtained offer new insights into *R. conorii* survival within the eschar site.

Results

The transcriptome profile of *R. conorii* is highly conserved among different eschars

The bacterial RNA samples purified from 8 eschars derived from 7 individual patients (Table 1) were subjected to whole-genome-wide transcript expression profile analysis and compared to the transcriptome of *R. conorii* grown *in vitro* and used as reference. As illustrated by Figure 1, human skin biopsies of MSF patients are histologically dominated by severe cutaneous necrosis with coagulative necrosis of the epidermis. Such specimens contain *R. conorii*, the bacteria being mostly found in necrotic areas associated with inflammatory cells. Given the scarce amount of available material, only one sample was hybridized twice. Hierarchical clustering of the signal intensities of the individual transcripts in both groups evidenced a high similarity of transcript expression patterns among eschars or infected cell monolayers, respectively. Measurements derived from similarity matrix indicated that largest distance among Vero cells profiles was 0.056 whereas a distance of 0.188 was found for the different eschars (Fig. 2). To accurately assess variations in gene expression within the group ⟨⟨eschar⟩⟩, the phylogenic analysis of the different clinical isolates was achieved with the multispacer typing (MST) based on the sequences of 3 variable intergenic spacers, namely *dksA-xerC*, *mppA-purC*, *rpmE-tRNA(fMet)* [12]. By combining the genotypes obtained from these three intergenic spacers, only 3 MST genotypes were obtained (Table 1). It is interesting to note that the two *R. conorii* isolates for which the transcriptome profiles appeared divergent from the main cluster (samples G and H) exhibited distinct MST genotypes.

The microarray results were confirmed by qRT-PCR for a subset of 16 targets. When comparing both methods, a high correlation coefficient ($R^2 = 0.934$) was observed (Fig. 3).

General overview

When compared with bacteria grown in Vero cell monolayers, the *R. conorii* gene transcripts in eschars were mainly found to be down-regulated. Of the 5,098 probes represented on our microarrays, 211 transcripts representing 15.4% (211/1374) of the *R. conorii* ORFs were differentially expressed ($P<0.05$). Of these, 180 genes were down-regulated two-fold or greater *in vivo* (supplementary Table S2) while only 31 genes exhibited an increased expression (Table 2). In the *R. conorii* genome, the size of the 1,374 annotated ORFs is ranging from 120 to 6,063 nucleotides (nt) with a mean value of 804 nt [13]. We noticed that 87% (27/31) of up-regulated genes within eschars have a size ranging from 153 nt to 723 nt, being thus significantly smaller than the median value.

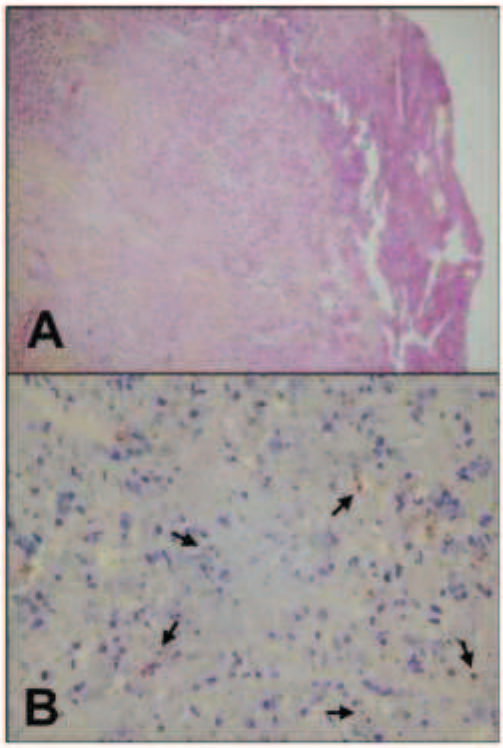


Figure 1. Inoculation eschar from a patient with MSF. (A) Histologic examination of inoculation eschar from a patient with MSF showing numerous dermal inflammatory infiltrates mainly composed of polymorphonuclear leukocytes with large necrotic areas and coagulative necrosis of the epidermis (hematoxylin-eosin-saffron, original magnification ×100). (B) Immunohistochemical detection of *R. conorii* in the inoculation eschar. Anti-*R. conorii* antibodies were detected using biotinylated secondary antibody, followed by avidin-peroxidase color development. The bacteria thus stained in reddish/brown, and indicated by the arrows, appear located between the necrotic inflammatory cells present in the dermis (original magnification ×400).
doi:10.1371/journal.pone.0003681.g001

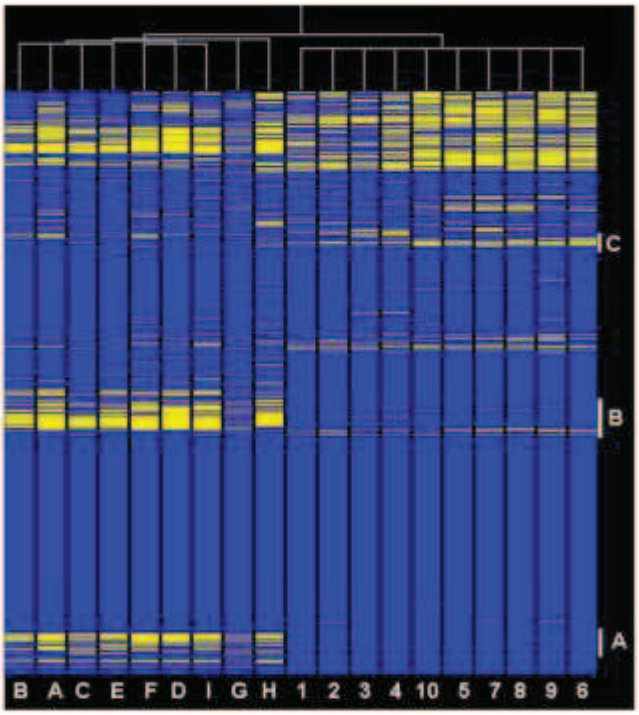


Figure 2. Two-way clustering analysis of *R. conorii* transcriptomic patterns. The figure shows a clear separation of two clusters containing 8 eschars recovered from different infectious episodes and 10 slides corresponding to bacteria grown in Vero cells and used as control, respectively. Each probe set is represented by a single row of colored boxes and each sample (eschar or control) correspond to a single column. The blue areas correspond to genes showing high or medium expression whereas yellow bars indicates genes poorly or not expressed. The dendrogram (white lines) on the top of the figure represents the similarity matrix of probe sets. Among clusters of genes allowing discriminating between the 2 tested conditions, A contains 70 genes involved mainly in different metabolic functions (transporters, DNA repair enzyme *mutL*, metabolic enzymes and numerous ribosomal proteins), cluster B contains an important number of genes involved in energy production (ATP synthesis) as well as genes involved in stress-response (*uvrA* and C and *htpG*), cell division and some virulence factors (*virB4* and *virB10*) and antibiotic resistance determinants. Cluster C contains several hem factors.
doi:10.1371/journal.pone.0003681.g002

We also observed that 21 of these genes, i.e. 67.7% are lacking or highly degraded in the typhus group (TG) rickettsiae. In contrast, the mean size of down-regulated genes was of 1,323 nt and 75% of them belong to the core gene of rickettsiae [14].

Evidence for rickettsial growth arrest within eschars

The ORFs of *R. conorii* whose expression was significantly altered were classified into functional categories according to the Cluster of Orthologous Gene (COG) classification, as defined by the COGs database [15]. As illustrated Figure 4, the most down-regulated COGs were translation (J), cell wall and membrane biogenesis (M), intracellular trafficking and secretion (U) as well as energy production and conversion (C). The reduced expression of 24 genes encoding ribosomal proteins indicates that, within eschars, rickettsiae are reducing their translational capacity (supplementary Table S2). Eighteen of these genes (from RC0981 to RC1007) are grouped together on the rickettsial chromosome and are organized as a ribosomal protein gene cluster. Consistent with a bacterial growth arrest, we also noticed a dramatic decrease in the expression of cell wall components including several Sca family proteins (Sca0, Sca1, Sca4, Sca5 and Sca10). Following 23SrRNA (105.5-fold decrease), the most down-regulated genes are those encoding the rickettsial outer membrane protein (rOmpB otherwise called Sca5; 98.5-fold decrease) and the rickettsial adhesin Adr2 (79.7-fold decrease). While down-regulated in a lower extend (fold-change mean of 5.4), many of the variable genes from the U category, including VirB, SecE, SecF and SecY subunits, encode proteins associated with the cell membrane. We also noticed the down-regulation of genes involved in energy production that should also contribute to bacterial replication arrest. Finally, 18% of down-regulated genes have no clearly defined function (S and R) or are not assigned to any functional categories.

R. conorii response to the host attack

Among the 31 rickettsial genes up-regulated in eschars, several are plausibly involved to escape host response. They include genes coding for proteins involved in DNA repair and modification as RC0550 (RecB exonuclease), RC1204 (tmRNA-binding protein), RC1050 (transposase) and RC1125 (helicase). Bacteria also face to oxidative stress, as indicated by the increase expression of *phbC* (poly-beta-hydroxybutyrate polymerase), *trxB2* (thioredoxin) and *grxC1* (glutaredoxin). An enhanced tolerance to osmotic stress could be provided through the increased expression of *proP9* (proline-betaine transporter) and *nuoL3* (NADH dehydrogenase chain I). Adaptation of *R. conorii* within eschars promoted changes in cell-wall related proteins as glycosyltransferase and cephalosporin hydroxylase (*cmcI*) and was accompanied with the up-regulation of several ATP binding cassette (ABC) transporters (RC0500, *mdlB*, *atm1*) and of two proteins that belong to the KAP NTPases, namely the NACHT NTPase and the putative AAA+ family ATPase, a new family characterized by the presence of transmembrane segments inserted into the P-loop NTPase domain [16]. We also noticed the over-expression of two genes encoding tetratricopeptide repeat (TPR)-containing proteins, suggesting the importance of this motif for protein-protein interactions between *R. conorii* and its host during infection, as previously evoked [17]. Apart from these, we also observed an increased expression of RC1370 and RC1298 genes that encode a prophage antirepressor and lysozyme, respectively. Finally, 11 out of the 31 up-regulated genes encode proteins with unknown function.

Discussion

Infection with *R. conorii* usually occurs following infected tick bite [2] and the first step of this host-pathogen interaction takes place at the inoculation site where anti-rickettsial immunity is enhanced [7]. The presence of an eschar in rickettsial disease is generally linked with a milder disease as more severe rickettsiosis, namely the Rocky Mountain Spotted Fever (RMSF) and epidemic typhus do not exhibit inoculation eschars [1,2,18]. This suggests that eschar corresponds to the front line of human host defense against rickettsia diffusion, a point consistent with the fact that the highest bacteremia were detected in patients suffering from RMSF and typhus [19]. Accordingly, the factors involved in the bacterial survival strategy can be considered as crucial actors of *R. conorii* pathogenesis. We thus examined the expression patterns of *R. conorii* in such environment, using an approach successfully applied for the analysis of transcriptional profile of bacteria within infected epithelial monolayers [11].

The human skin biopsies being rare and precious, the optimal conditions for total RNA extraction were first assayed on eschars experimentally induced through the intradermal injection of *R. conorii* in rabbits (not shown). Starting material used in this study

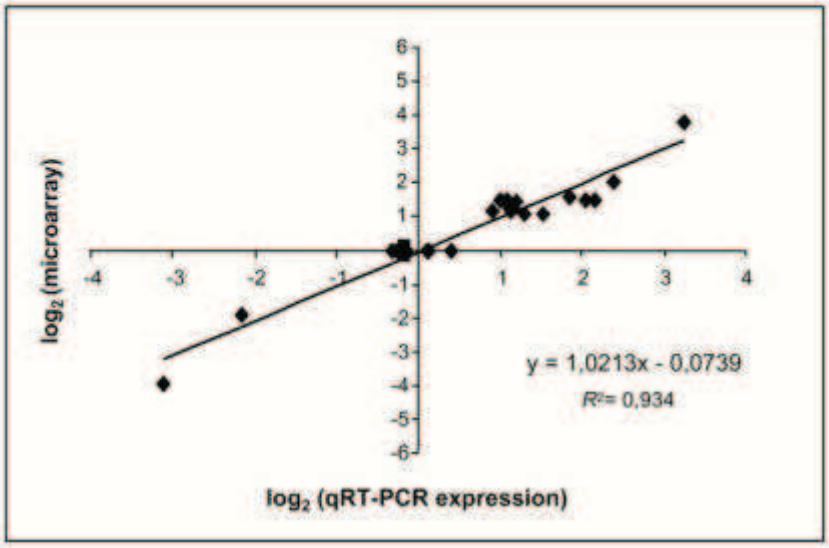


Figure 3. Validation of microarray-based expression profiles by qRT-PCR. The relative transcriptional levels for 16 genes were determined by real-time qRT-PCR using cDNA as template. Following normalization of data based on values measured with non regulated transcripts, the real-time qPCR $\log_2$ values were plotted against the microarray $\log_2$ values. The correlation coefficient (R^2) for comparison of the two datasets is of 0.934.
doi:10.1371/journal.pone.0003681.g003

was collected at different time points after the tick bite and stored at −80°C for variable periods of time (up to 3 years) before processing. We predicted that this factor would interfere with the identification of differences in gene expression because of biological variations. In addition, the pattern of gene expression may also vary depending on the genetic background of infecting strains. Indeed, in microarray-based experiments, a lower hybridization can result either from a reduced amount of transcripts or from divergence in the sequence of the gene [20]. Thus, insertion or deletion events in the genome of clinical strains tested could affect the apparent transcript abundance measured by using a microarray designed from the genome sequence of the Malish strain (seven) of *R. conorii* [13]. Here, three *R. conorii* strains corresponding to cases issued from three different geographical sites were identified. However, analysis of rickettsial transcript expression patterns from the 8 human eschars included in this study yielded reproducible results and only minor variations were observed between biopsy specimens. Finally, results obtained by qRT-PCR indicated that reliable microarray hybridizations can be achieved with rickettsial RNA extracted from multiple and independently obtained eschars.

This work showed that within eschars, the *R. conorii* transcripts were mainly down-regulated compared to bacteria internalized in Vero cells. The most significantly repressed genes are those of the translation machinery. This observation is consistent with previous analysis of the transcriptional changes displayed by 19 different bacterial pathogens upon eukaryotic cell infection [8]. Within these conditions, a general decline of genes involved in general metabolism associated with bacterial growth (translation, transcription, cell wall biogenesis, energy production, transport of carbohydrates, amino acids and nucleotides) was observed. Such a global expression decrease resembles that depicted for *Bacillus subtilis* [21], *Escherichia coli* [22], *Corynebacterium glutamicum* [23], and *Staphylococcus aureus* [24] after inducing the stringent response [25]. The recent analysis of inflammatory and immune mediators present in skin-biopsy samples of patients suffering from MSF evidenced the production of some enzymes including inducible nitric oxide synthase and indolcamine-2,3-dioxygenase, that should contribute to the bacterial growth arrest [7]. However, we observed that in some cases, the immune and inflammatory host response was therefore not strong enough to eradicate all infecting *R. conorii* as the bacterium was cultivable from 3 (62.5%) of the eschar biopsy specimens.

Analysis of obtained data highlighted some of the mechanisms displayed by the rickettsiae to counter the damage and survive within the host cells. Genes encoding proteins involved in genome repair and allowing bacteria to cope with DNA-damaging agents are probably critical in this situation and were up-regulated. DNA repair is a fundamental process used by pathogenic bacteria as one of the defense mechanisms that allow them to survive in their hosts [26]. As described for *Helicobacter pylori* [27], the DNA lesions could result from a cellular oxidative stress, a point consistent with the increased rickettsial defenses against reactive oxygen species. Noteworthy is that production of bactericidal reactive oxygen species is one of the key methods by which mammalian infected cells efficiently kill bacteria [28]. Within eschars, *R. conorii* also deal with osmotic stress as indicated by the up-regulation of proline-betaine transporter and of NADH dehydrogenase I. Proline and betaine are two osmoprotectants accumulated through enhanced uptake rather than synthesis by Gram negative bacteria to overcome the inhibitory effects of hyperosmolarity [29]. While the *nuo* genes of rickettsiae are more closely related to mitochondria than to any other studied microbe [30], enhancement of NADH dehydrogenase expression could correspond to another strategy of osmoadaptation evolved to achieve salt tolerance. *R. conorii* also displayed another hallmark feature of pathogens interacting with host cells, namely the phenotypic changes in the composition of several membrane proteins [8]. It is well established that antigenic variation is an important mechanism that allows pathogens to escape for immunity. In this respect, we noticed an important down-regulation of genes encoding for the Sca family proteins among which Sca0 (rOmpA) and Sca5 (rOmpB) which are two major rickettsial antigenic determinants [31]. Antigenic variation could also be related to the post-translational modifications of proteins. These changes could be afforded by CmcI, a protein recently classified as a methyltrans-

Table 2. Thirty one genes of *R. conorii* up-reguled within eschars.

ORFs	Genes	Description	Fold-Ch	Size (nt)
RC0550		RecB family exonuclease	13,685	2,517
RC0828**		**Unknown**	**6,763**	**183**
RC0500		**ABC transporter ATP-binding protein**	**4,019**	**723**
RC1370*		**Prophage antirepressor (SPLIT GENE)**	**3,790**	**207**
RC0769**		**Unknown**	**3,661**	**156**
RC1137	*phbC*	**Poly-beta-hydroxybutyrate polymerase (SPLIT GENE)**	**3,628**	**264**
RC1298**		**Lysozyme (FRAGMENT)**	**3,405**	**204**
RC0350**		**Unknown**	**3,225**	**240**
RC1204	*smpA*	**tmRNA-binding protein**	**3,068**	**474**
RC0299*	*mdlB*	**ABC-type multidrug transport system, ATPase and permease components (SPLIT GENE)**	**2,986**	**273**
RC1299**		**Unknown**	**2,972**	**165**
RC0921**		**Unknown**	**2,825**	**153**
RC1349**		**Unknown**	**2,814**	**201**
RC0637	*trxB2*	**Thioredoxin reductase [EC:1.6.4.5]**	**2,785**	**1,023**
RC0267	*grxC1*	**Glutaredoxin**	**2,780**	**309**
RC0461*		**Glycosyltransferase [EC:2.4.1.-], two domains**	**2,708**	**1,815**
RC1050**		**Transposase (FRAGMENT)**	**2,655**	**171**
RC1125*		**Superfamily I DNA and RNA helicases (SPLIT GENE)**	**2,617**	**192**
RC0378	*nuoL3*	**NADH dehydrogenase I chain L [EC:1.6.5.3]**	**2,507**	**1,518**
RC1251		**Unknown**	**2,390**	**504**
RC0221**	*cmcI*	**Cephalosporin hydroxylase (FRAGMENT)**	**2,342**	**321**
RC0914		**Tetratricopeptide repeat-containing protein**	**2,271**	**264**
RC0957*		**Tetratricopeptide repeat-containing protein (FRAGMENT)**	**2,263**	**228**
RC1155*		**Unknown (SPLIT GENE)**	**2,256**	**348**
RC0209*		**Unknown**	**2,225**	**282**
RC0795**		**NACHT family NTPase (FRAGMENT)**	**2,146**	**213**
RC0890*	*proP9*	Proline/betaine transporter (FRAGMENT)	2,102	515
RC0269	*atm1*	Multidrug resistance protein Atm1 (SPLIT GENE)	2,085	201
RC0060**		**Unknown**	**2,067**	**492**
RC1144**		**Unknown**	**2,066**	**231**
RC1192**		**Putative AAA+ superfamily ATPase (SPLIT GENE)**	**2,061**	**183**

Genes absent (**) or degraded (*) in the TG.
doi:10.1371/journal.pone.0003681.t002

ferase based on structural evidences [32], or by a glycosyltransferase. Because most glycoproteins were associated with virulence factors in bacterial pathogens [33], these events could contribute to differences in both virulence and antigenicity of *R. conorii in vivo*.

Clearly, besides evasion of host defense, rickettsiae also exhibited virulence determinants *in vivo*. Thus, the export of virulence proteins in the host cell cytoplasm could be achieved through the increased expression of ABC transporters that function as type I secretion system in Gram negative bacteria [34]. The way in which rickettsiae use the KAP NTPases in the intracellular host cell environments has not yet been investigated but several members of this family, namely the AAA+ ATPases were described to promote virulence of other bacterial pathogens [35]. A role for the α-superhelical structure domain of KAP NTPases has been evoked in protein-protein interactions [16]. Interestingly, such interactions can also be mediated by the TPR [36], another structural motif present within two *R. conorii* proteins up-regulated *in vivo*. Here again, the functional role of these ubiquitous domains was not deciphered. However, TPR repeat regions have been implicated in the ability of *L. pneumophila* to efficiently establish infection and/or to manipulate host cell trafficking event [37]. The whole picture of this host-pathogen interaction within the first step of infection is summarized Figure 5.

A more general view of obtained results showed that, to ensure microbial fitness and survival in the lethal host environment, *R. conorii* mainly promoted the transcription of small size genes. While a high proportion of them codes for proteins with unknown functions, those with functional attributes including *phbC*, *mdlB* and *proP9* have already been associated with pathogenicity in other bacteria. Another interesting feature is that up-regulated genes are mostly exclusive for SFG rickettsiae, a finding consistent with the fact that the inoculation site corresponds to environmental conditions not encountered by TG bacteria that are transmitted to human by exposure to feces of infected lice or fleas [1].

Recent studies showed that an understanding of the basic mechanisms of adaptation of rickettsiae under various environ-

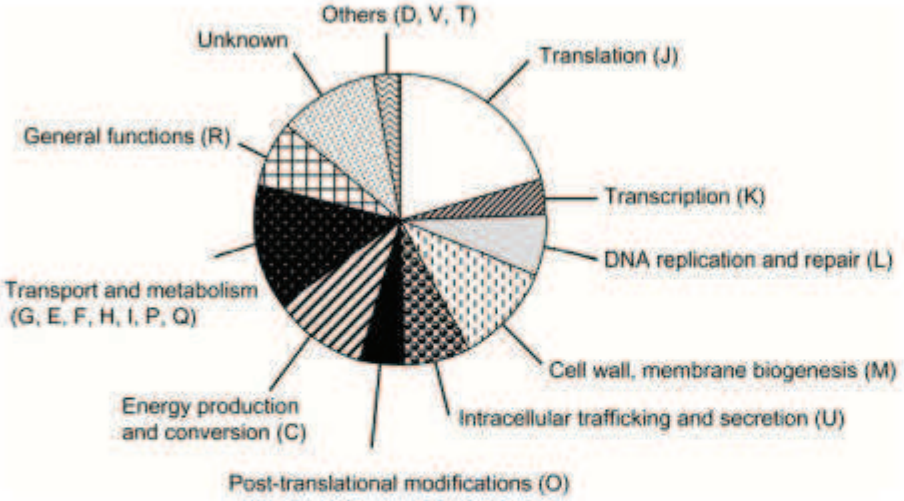


Figure 4. Distribution of *R. conorii* ORFs down-regulated during infection. Pie chart showing the percentages of transcripts down-regulated in eschars and classified according to their respective functional categories. COG category legend is as follow: Cell cycle control, mitosis and meiosis (D); Defense mechanisms (V); Signal transduction mechanisms (T); Transport and metabolism of carbohydrates (G), amino acids (E), nucleotides (F), coenzymes (H), lipids (I), inorganic ions (P), and of secondary metabolites (Q).
doi:10.1371/journal.pone.0003681.g004

ments, including nutrient deprived medium [11] and temperature changes [38] can be gained by determining their global transcriptome profile by microarrays. As shown here, this approach also offers the opportunity to characterize strategies displayed by the bacteria for *in vivo* survival. A new picture of *Rickettsia* pathogenicity, still poorly known, has thus emerged. A better knowledge of such an host-pathogen interaction would offer the opportunity to identify future therapeutic targets usefull for the prevention or the treatment of rickettsiosis.

Materials and Methods

Human subjects and sample collection

From June 2004 through August 2006, cutaneous biopsy specimens were obtained from seven patients in Marseille with eschar-associated illnesses, and who were suspected of suffering from MSF on the basis of initial clinical findings or laboratory analysis. These biopsy specimens, excised by using a scalpel, were dissected in two pieces. For RNA extraction purpose, the tissues

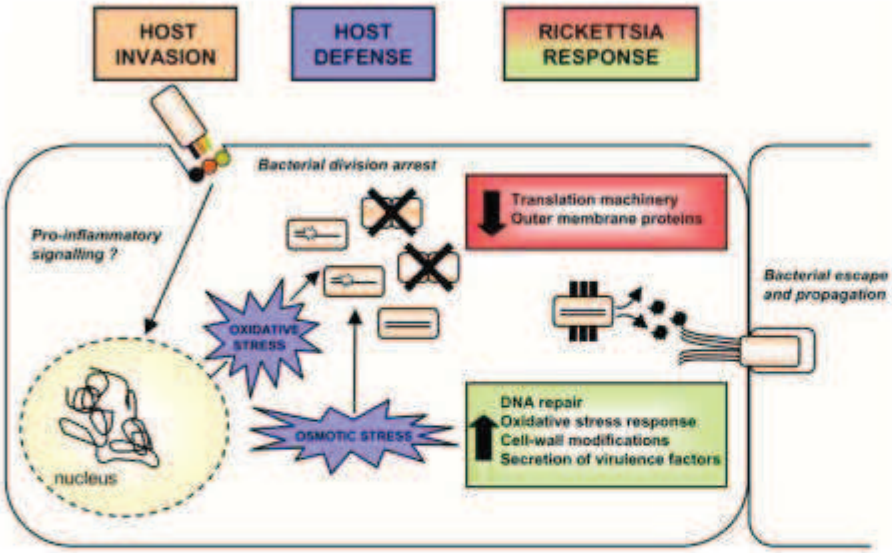


Figure 5. Schematic representation of the interaction between *R. conorii* and its host cell at the site of inoculation. Intracellular uptake of *R. conorii* by induced phagocytosis [46] may activate signalling pathways in the host cells which in turn display several mechanisms to eliminate invading bacteria among which oxidative and osmotic stress. As shown by the work presented in this paper, these events induce a strong down-regulation of *R. conorii* transcripts, mainly of those involved in bacterial replication and classified within translation and cell wall membrane COGs. These pathogenic bacteria therefore evade the host defense through the up-regulation of several factors counteracting DNA damages and through variations of dominant surface antigens allowing to avoid host recognition. The secretion of virulence determinants is also likely to favour survival and colonization of the host.
doi:10.1371/journal.pone.0003681.g005

were collected in sterile Eppendorf tubes containing RNA*later* Stabilization Reagent (Qiagen, Courtaboeuf, France) refrigerated at 4°C overnight before storage at −80°C. The second half of specimens was processed for bacterial culture and PCR-based molecular diagnosis, histopathology as well as immunohistochemical staining after fixation in ethanol. Selected clinical variables of the individual patients are shown in Table 1. These experiments were carried out with both the approval from the local ethics committee (IFR 48 ethics committee) and the written consent of informed patients.

Histologic analysis and immunohistochemical detection of *R. conorii*

Formalin-fixed, paraffin-embedded skin biopsy specimens of the inoculation eschars were cut (3 µm thickness) and stained with hematoxylin-eosin-saffron by routine staining methods. Serial sections of each tissue specimen were also obtained for immunohistochemical investigations. The immunohistochemical analysis was performed by the indirect immunoperoxidase method has described elsewhere [39] and using a polyclonal rabbit antibody against *R. conorii* as primary antibody. Briefly, after deparaffinization, each tissue section was incubated with the polyclonal anti-*R. conorii* antibody (1:2000) and immunodetection was performed with biotinylated immunoglobulins, followed by peroxidase-labeled streptavidin (HistoStain plus kit, Zymed, Montrouge, France) with amino-ethyl-carbazole as substrate. The slides were counterstained with Mayer hematoxylin for 10 min. Images were acquired with a Zeiss Axioskop microscope coupled with a Nikon Coolpix 4500 digital camera.

Multispacer genotyping (MST) of *R. conorii* clinical isolates

Total genomic DNA was extracted from skin-biopsy specimens using the QIAamp Tissue kit (Qiagen), as described by the manufacturer. Amplification of the three highly variable intergenic spacers, *dksA-xerC*, *mppA-purC*, and *rpmE-tRNA*^{fMet} was achieved using previously described primers [12] and HotStarTaq DNA polymerase (Qiagen). All primers were obtained from Eurogentec (Seraing, Belgium). PCR was carried out in a PTC-200 automated thermal cycler (MJ Research, Waltham, Mass.) under the following conditions: an initial 15 min-step at 95°C was followed by 39 cycles of denaturation for 30 s at 94°C, annealing for 30 s at 54°C, and extension for 1 min at 72°C. Final amplification was completed by holding the reaction mixture for 5 min at 72°C to allow complete extension of the PCR products. Following purification (MultiScreen PCR filter plate, Millipore, Saint-Quentin en Yvelines, France), amplicons were sequenced in both directions using the BigDye 1.1 chemistry (Applied Biosystems) on an ABI 3130XL automated sequencer (Applied Biosystems) as described by the manufacturer. To avoid contamination, no positive control was used. Sterile water was used as a negative control in each PCR assay. Sequences from each DNA sample were checked twice in both directions to ensure the reliability of the MST method, then assembled and edited with the Sequencher 4.7 program (GeneCode, Ann Arbor, Mich.). For the phylogenetic analysis, the sequences of three spacers were concatenated. Multiple alignments of the concatenated spacer sequences were carried out using the CLUSTALW software [40]. Phylogenetic relationships were obtained using the neighbor-joining and maximum parsimony methods within the MEGA 4.1 software [41].

Microarray design

ArrayDesigner™ (Premier Biosoft) was used to generate an initial set of probes covering the whole genome of *R. conorii* strain Malish 7 (NC_003103), showing specific physico-chemical properties: (i) 71±12° C target Tm, (ii) 40–60 bp probe length, (iii) <−5.0 kcal/mol for hairpins, (iv) <−8.0 kcal/mol for self-dimers, and (v) dinucleotide repeats shorter than 5 bp. Candidate probes were tested for specificity against the aforementioned *R. conorii* genome using Olicheck [42]. To avoid cross-hybridization with host cell nucleic acids, the resulting set of probes was subjected to Blast analysis against the *Homo sapiens* and *Pan troglodytes* (by default of the African green monkey genome, Vero cells). Probes with >18 consecutive nt matches were excluded. Preceding steps yielded a final oligonucleotide set of 5,098 probes resulting to a final coverage of 97% for ORFs and 63% for inter-ORFs (considering fragment with length >149 bp which is the median size of inter-ORFs). To minimize steric hindrance, all probes <60 nt in length were poly(T)-tailed to reach an overall length of 60 nt, following Agilent's recommendations. Microarrays (2×11K format) were manufactured by *in situ* synthesis SurePrint technology (Agilent Technologies, Palo Alto, CA). All specific oligonucleotides as well as Agilent's control spots were printed in duplicate.

RNA isolation and purification from eschars

Total RNA extraction and purification from the eschar biopsies was carried out using the RNeasy kit (Qiagen) with some modifications. Briefly, the tissue samples removed from RNA*later* were rapidly decontaminated with iodated alcohol. After 2 washings in RNase-free water, 20 mg of tissue excised in small pieces were homogenized in RLT solution with tungsten beads and using the Mixer Mill MM300 (Qiagen). The resulting homogenate was then incubated at 55°C for 10 min with proteinase K (200 µg/ml) and centrifuged for 3 min at 10,000×*g*. Supernatant containing the total RNA fraction was supplemented with ethanol and purified onto RNeasy columns according to the manufacturer's instructions. The amount and quality of obtained RNA were determined with the microfluidic-based platform (Agilent 2100 Bioanalyzer) and using the RNA 6000 Nano Labchip kit (Agilent). In the electropherograms obtained with total RNA, the prokaryotic fraction was not always detected because its low abundance and the profiles of eukaryotic peaks were used as indicators for the integrity of both RNA populations. Estimated amount of total RNA extracted from 20 mg eschar was generally around 10 µg, including 90% of eukaryotic RNAs. These contaminants were removed using the MICROBEnrich procedure (Ambion, Applied Biosystems, Courtaboeuf, France) and prokaryotic cDNA was synthesized as described [10] using the M-MLV Reverse Transcriptase (Invitrogen, Cergy-Pontoise, France).

RNA isolation and purification from infected Vero cells

R. conorii strain Malish 7 (ATCC, VR613) grown in Vero cells for 48 h at 32°C were lysed in TE buffer (10 mM Tris/HCl, 1 mM EDTA pH 7.0) supplemented with lysozyme (10 mg/ml) for 10 minutes at room temperature. Total RNA was then extracted and purified using RNeasy-Midi columns (Qiagen), as previously described [11]. Two batches bacteria grown separately (biological replicates) were collected for this study. We then applied the same procedure as for eschar samples for purification, amplification, labelling and hybridization of *R. conorii* cDNA.

Expression microarrays and analysis

Twenty nanograms of cDNA were amplified using the GenomiPhi DNA amplification kit (Amersham Biosciences, Uppsala, Sweden) and labeled with Cy3-dCTP or Cy5-dCTP dyes (Amersham Biosciences) using the BioPrime DNA labeling System (Invitrogen, Cergy-Pontoise, France). Following purifica-

tion with GFX columns (Amersham Biosciences), the levels of Cy3-dCTP and Cy5-dCTP incorporation were quantified by absorbance measurement at 550 nm and 650 nm, respectively. Hybridizations were performed for 17 h at 60°C in dedicated micro-chambers with 75 pmol of both control or eschar samples. Stringent washings were then performed according to manufacturer's instructions. Slides dried by 30 sec washing with acetonitrile were scanned using 100% Photon Multiplier Tube power for both wavelengths using Agilent scanner (Agilent Technologies, CA, USA) and extracted using Feature Extraction™ software (version 8, Agilent). Local background subtracted signals were corrected for unequal dye incorporation or unequal load of the labeled product and normalized using GeneSpring (Agilent). The algorithm consisted of a rank consistency filter and a curve fit using the default LOWESS (locally weighted linear regression) method. Data consisting of 10 independent control conditions hybridized against 10 independent patient eschar experiments were expressed as Log10 ratios and analyzed using GeneSpring 7.3 (Agilent). Statistical significance of differentially expressed genes was identified by variance analysis (ANOVA) [43,44] performed using GeneSpring, including the Benjamini and Hochberg false discovery rate correction (5%). Expression microarrays (normalized data) were clustered by a hierarchical clustering algorithm by using an average linkage method in GeneSpring. The expression values for a gene across all samples were standardized to have mean of 0 and standard deviation of 1 by linear transformation. To determine the amount of detectable genes, the expression values were averaged for transcripts mapped by 2 or more probes. A cut-off value defined as 2×standard deviation obtained for background intensities was then applied [45].

Microarray data accession numbers

The data have been deposited in NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/). The GEO accession numbers are GPL7040 for the platform of the microarray and GSE12130 for the experimental data set.

Real-time quantitative PCR

Validation of microarray data was achieved using cDNA synthesized from eschar specimens A/B, C and E and from *R. conorii* grown in Vero cells used as reference. Real-time quantitative RT-PCR (qRT-PCR) was performed on the SmartCycler system (Cepheid) together with the QuantiTec Probe PCR kit (Qiagen) or SybrGreen DNA Fast Start kit (Roche Diagnostics, Basel, Switzerland), as indicated. Gene-specific primers are listed in the Supplementary Table S1. The values obtained for the 3 non-differentially expressed genes (*trxB1*, *glyQ* and *dnaK*) were used to normalize all data. The fold change (FC) in expression of the target genes relative to the 3 unregulated genes was determined as follows: $FC = 2^{-\Delta\Delta Ct}$ where $\Delta\Delta Ct = (\text{Mean-Ct}_{\text{target}} - \text{Mean-Ct}_{\text{control/eschars}}) - (\text{Mean-Ct}_{\text{target}} - \text{Mean-Ct}_{\text{control/reference}})$. Ct values were defined as the cycle numbers at which the fluorescence signals were detected.

Supporting Information

Table S1 List of primers and conditions for qRT-PCR assays
Found at: doi:10.1371/journal.pone.0003681.s001 (0.04 MB XLS)

Table S2 R. conorii genes down-regulated within eschars
Found at: doi:10.1371/journal.pone.0003681.s002 (0.08 MB XLS)

Author Contributions

Conceived and designed the experiments: PR CR JS DR. Performed the experiments: CR QL WL HL. Analyzed the data: PR QL AH WL HL PF. Contributed reagents/materials/analysis tools: JS AH DR. Wrote the paper: PR PF DR.

References

1. Raoult D, Roux V (1997) Rickettsioses as paradigms of new or emerging infectious diseases. Clin Microbiol Rev 10: 694–719.
2. Parola P, Paddock CD, Raoult D (2005) Tick-borne rickettsioses around the world: emerging diseases challenging old concepts. Clin Microbiol Rev 18: 719–756.
3. Montenegro MR, Mansueto S, Hegarty BC, Walker DH (1983) The histology of "taches noires" of boutonneuse fever and demonstration of *Rickettsia conorii* in them by immunofluorescence. Virchows Arch A Pathol Anat Histopathol 400: 309–317.
4. Raoult D, de Micco C, Gallais H, Toga M (1984) Laboratory diagnosis of Mediterranean spotted fever by immunofluorescent demonstration of *Rickettsia conorii* in cutaneous lesions. J Infect Dis 150: 145–148.
5. Dujella J, Morovic M, Dzelalija B, Gveric M, Novakovic S (1991) Histopathology and immunopathology of skin biopsy specimens in Mediterranean spotted fever. Acta Virol 35: 566–572.
6. Walker DH, Occhino C, Tringali GR, Di Rosa S, Mansueto S (1988) Pathogenesis of rickettsial eschars: the tache noire of boutonneuse fever. Hum Pathol 19: 1449–1454.
7. de Sousa R, Ismail N, Nobrega SD, França A, Amaro M, et al. (2007) Intralesional expression of mRNA of interferon- gamma , tumor necrosis factor-alpha , interleukin-10, nitric oxide synthase, indoleamine-2,3-dioxygenase, and RANTES is a major immune effector in Mediterranean spotted fever rickettsiosis. J Infect Dis 196: 770–781.
8. La MV, Raoult D, Renesto P (2008) Regulation of whole bacterial pathogen transcription within infected hosts. FEMS Microbiol Rev 32: 440–460.
9. Waddell SJ, Butcher PD, Stoker NG (2007) RNA profiling in host-pathogen interactions. Curr Opin Microbiol 10: 297–302.
10. Francois P, Garzoni C, Bento M, Schrenzel J (2007) Comparison of amplification methods for transcriptomic analyses of low abundance prokaryotic RNA sources. J Microbiol Methods 68: 385–391.
11. La MV, Francois P, Rovery C, Robineau S, Barbry P, et al. (2007) Development of a method for recovering rickettsial RNA from infected cells to analyze gene expression profiling of obligate intracellular bacteria. J Microbiol Methods 71: 292–297.
12. Fournier PE, Zhu Y, Ogata H, Raoult D (2004) Use of highly variable intergenic spacer sequences for multispacer typing of *Rickettsia conorii* strains. J Clin Microbiol 42: 5757–5766.
13. Ogata H, Audic S, Renesto-Audiffren P, Fournier PE, Barbe V, et al. (2001) Mechanisms of evolution in *Rickettsia conorii* and *R. prowazekii*. Science 293: 2093–2098.
14. Blanc G, Ogata H, Robert C, Audic S, Suhre K, et al. (2007) Reductive genome evolution from the mother of *Rickettsia*. PLoS Genet 3: e14.
15. Tatusov RL, Galperin MY, Natale DA, Koonin EV (2000) The COG database: a tool for genome-scale analysis of protein functions and evolution 225. Nucleic Acids Res 28: 33–36.
16. Aravind L, Iyer LM, Leipe DD, Koonin EV (2004) A novel family of P-loop NTPases with an unusual phyletic distribution and transmembrane segments inserted within the NTPase domain. Genome Biology 5: R30.
17. Ogata H, Robert C, Audic S, Robineau S, Blanc G, et al. (2005) *Rickettsia felis*, from culture to genome sequencing. Ann NY Acad Sci 1063: 26–34.
18. Tbourner AR, Walker DH, Petri WAJ (1998) Rocky mountain spotted fever. Clin Infect Dis 27: 1353–1359.
19. Tzianabos T, Anderson BE, McDade JE (1989) Detection of *Rickettsia rickettsii* DNA in clinical specimens by using polymerase chain reaction technology. J Clin Microbiol 27: 2866–2868.
20. Le Gall T, Darlu P, Escobar-Páramo P, Picard B, Denamur E (2005) Selection-driven transcriptome polymorphism in *Escherichia coli/Shigella* species. Genome Res 15: 260–268.
21. Eymann C, Homuth G, Scharf C, Hecker M (2002) *Bacillus subtilis* functional genomics: global characterization of the stringent response by proteome and transcriptome analysis. J Bacteriol 184: 2500–2520.
22. Chang D-E, Smalley DJ, Conway T (2002) Gene expression profiling of *Escherichia coli* growth transitions: an expanded stringent response model. Mol Microbiol 45: 289–306.
23. Brockmann-Gretza O, Kalinowski J (2006) Global gene expression during stringent response in Corynebacterium glutamicum in presence and absence of the rel gene encoding (p)ppGpp synthase. BMC Genomics 7: 230.
24. Garzoni C, Francois P, Huyghe A, Couzinet S, Tapparel C, et al. (2007) A global view of *Staphylococcus aureus* whole genome expression upon internalization in human epithelial cells. BMC Genomics 8: 171.
25. Cashel M, Gentry DR, Hernandez VJ, Vinella D (1996) The stringent response. In: Neidhardt FC, Curtis III R, Ingraham JL, Lin ECC, Low KB, et al., eds.

Escherichia coli and *Salmonella*: Cellular and Molecular Biology. Washington, D.C.: ASM Press. pp 1458–1496.
26. Heithoff DM, Conner CP, Mahan MJ (1997) Dissecting the biology of a pathogen during infection. Trends Microbiol 5: 509–513.
27. O'Rourke EJ, Chevalier C, Pinto AV, Thiberge JM, Ielpi L, et al. (2003) Pathogen DNA as target for host-generated oxidative stress: role for repair of bacterial DNA damage in *Helicobacter pylori* colonization. Proc Natl Acad Sci U S A 100: 2789–2794.
28. Radtke AL, O'Riordan MX (2006) Intracellular innate resistance to bacterial pathogens. Cell Microbiol 8: 1720–1729.
29. Csonka LN (1989) Physiological and genetic responses of bacteria to osmotic stress. Microbiol Rev 53: 121–147.
30. Andersson SG, Zomorodipour A, Andersson JO, Sicheritz-Ponten T, Alsmark UC, et al. (1998) The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. Nature 396: 133–140.
31. Teysseire N, Raoult D (1992) Comparison of Western immunoblotting and microimmunofluorescence for diagnosis of Mediterranean spotted fever. J Clin Microbiol 30: 455–460.
32. Oster LM, Lester DR, Terwisscha van Scheltinga A, Svenda M, van Lun M, G, et al. (2006) Insights into cephamycin biosynthesis: the crystal structure of CmcI from *Streptomyces clavuligerus*. J Biol Mol 358: 546–558.
33. Schmidt MA, Riley LW, Benz I (2003) Sweet new world: glycoproteins in bacterial pathogens. Trends Microbiol 11: 564–561.
34. Davidson AL, Maloney PC (2007) ABC transporters: how small machines do a big job. Trends Microbiol 15: 448–455.
35. Henderson B, Allan E, Coates ARM (2006) Stress wars: the direct role of host and bacterial molecular chaperones in bacterial infection. Infect Immun 74: 3693–3706.
36. D'Andrea LD, Regan L (2003) TPR proteins: the versatile helix. Trends Biochem Sci 28: 655–662.
37. Newton HJ, Sansom FM, Bennett-Wood V, Hartland EL (2006) Identification of *Legionella pneumophila*-specific genes by genomic subtractive hybridization with Legionella micdadei and identification of lpnE, a gene required for efficient host cell entry. Infect Immun 74: 1683–1691.
38. Dreher-Lesnick SM, Ceraul SM, Rahman MS, Azad AF (2008) Genome-wide screen for temperature-regulated genes of the obligate intracellular bacterium, *Rickettsia typhi*. BMC Microbiol 8: 61.
39. Lepidi H, Fournier PE, Raoult D (2000) Quantitative analysis of valvular lesions during *Bartonella* endocarditis. Am J Clin Pathol 114: 880–889.
40. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice 200. Nucleic Acids Res 22: 4673–4680.
41. Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: molecular evolutionary genetics analysis software. Bioinformatics 17: 1244–1245.
42. Charbonnier Y, Gentler B, Francois P, Bento M, Renzoni A, et al. (2005) A generic approach for the design of whole-genome oligoarrays, validated for genomotyping, deletion mapping and gene expression analysis on *Staphylococcus aureus*. BMC Genomics 6: 95.
43. Churchill GA (2004) Using ANOVA to analyze microarray data. Biotechniques 37: 173–175.
44. Scherl A, Francois P, Charbonnier Y, Deshusses JM, Koessler T, et al. (2006) Exploring Glycopeptide Resistance in *Staphylococcus aureus*: A combined proteomics and transcriptomics approach for the identification of resistance related markers. BMC Genomics 7: 296.
45. Scherl A, Francois P, Bento M, Deshusses JM, Charbonnier Y, et al. (2005) Correlation of proteomic and transcriptomic profiles of *Staphylococcus aureus* during the post-exponential phase of growth. J Microbiol Methods 60: 247–257.
46. Teysseire N, Boudier JA, Raoult D (1995) *Rickettsia conorii* entry into Vero cells. Infect Immun 63: 366–374.

Annexe 2

Analysis of the *Rickettsia africae* Genome Reveals that Virulence Acquisition in *Rickettsia species* may be Explained by Genome Reduction

Pierre-Edouard FOURNIER[1], Khalid El KARKOURI[1], Quentin LEROY[1], Catherine ROBERT[1], Bernadette GIUMELLI[1], Patricia RENESTO[1], Cristina SOCOLOVSCHI[1], Philippe PAROLA[1], Stéphane AUDIC[2] and Didier RAOULT [1]

BMC Genomics

2009 Apr 20;10:166

1 - Unité des Rickettsies, IFR 48 CNRS UMR 6020, Faculté de Médecine, Université de la Méditerranée, 27 Boulevard Jean Moulin, 13385 Marseille cedex 05, France

2 - Information Génomique et Structurale, CNRS UPR2589, Institut de Biologie structurale et Microbiologie, Marseille, France

BMC Genomics

BioMed Central

Research article

Open Access

Analysis of the *Rickettsia africae* genome reveals that virulence acquisition in *Rickettsia* species may be explained by genome reduction

Pierre-Edouard Fournier[1], Khalid El Karkouri[1], Quentin Leroy[1], Catherine Robert[1], Bernadette Giumelli[1], Patricia Renesto[1], Cristina Socolovschi[1], Philippe Parola[1], Stéphane Audic[2] and Didier Raoult*[1]

Address: [1]Unité des rickettsies, IFR 48 CNRS UMR 6020, Faculté de médecine, Université de la Méditerranée, 27 Boulevard Jean Moulin, 13385 Marseille cedex 05, France and [2]Information Génomique et Structurale, CNRS UPR2589, Institut de Biologie structurale et Microbiologie, Marseille, France

Email: Pierre-Edouard Fournier - Pierre-Edouard.Fournier@univmed.fr; Khalid El Karkouri - Khalid.Elkarkouri@univmed.fr; Quentin Leroy - Quentin.Leroy@univmed.fr; Catherine Robert - Catherine.Robert@univmed.fr; Bernadette Giumelli - Bernadette.Giumelli@univmed.fr; Patricia Renesto - prenesto@embl.fr; Cristina Socolovschi - Cristina.Socolovschi@univmed.fr; Philippe Parola - Philippe.Parola@univmed.fr; Stéphane Audic - Stephane.Audic@igs.cnrs-mrs.fr; Didier Raoult* - didier.raoult@gmail.com

* Corresponding author

Published: 20 April 2009

BMC Genomics 2009, **10**:166 doi:10.1186/1471-2164-10-166

Received: 7 November 2008
Accepted: 20 April 2009

This article is available from: http://www.biomedcentral.com/1471-2164/10/166

Abstract

Background: The *Rickettsia* genus includes 25 validated species, 17 of which are proven human pathogens. Among these, the pathogenicity varies greatly, from the highly virulent *R. prowazekii*, which causes epidemic typhus and kills its arthropod host, to the mild pathogen *R. africae*, the agent of African tick-bite fever, which does not affect the fitness of its tick vector.

Results: We evaluated the clonality of *R. africae* in 70 patients and 155 ticks, and determined its genome sequence, which comprises a circular chromosome of 1,278,540 bp including a *tra* operon and an unstable 12,377-bp plasmid. To study the genetic characteristics associated with virulence, we compared this species to *R. prowazekii*, *R. rickettsii* and *R. conorii*. *R. africae* and *R. prowazekii* have, respectively, the less and most decayed genomes. Eighteen genes are present only in *R. africae* including one with a putative protease domain upregulated at 37°C.

Conclusion: Based on these data, we speculate that a loss of regulatory genes causes an increase of virulence of rickettsial species in ticks and mammals. We also speculate that in *Rickettsia* species virulence is mostly associated with gene loss.

The genome sequence was deposited in GenBank under accession number [GenBank: NZ_AAUY01000001].

Background

Rickettsiae are obligate intracellular Gram-negative bacteria mostly associated to arthropods, some of which causing mild to severe diseases in humans. Pathogenic species are classified into two groups based on phylogenetic analyses [1]. The typhus group (TG) includes two *Rickettsia*

prowazekii (*R. prowazekii*) and *R. typhi*, and the spotted fever group (SFG) includes 15 pathogenic species and numerous species of unknown pathogenicity [2,3]. Two additional validated species, *R. bellii* and *R. canadensis*, and a variety of unvalidated species from insects or leeches are organized into the most outer outgroups of the genus *Rickettsia* [3-5]. The relatively low rate of lateral gene transfer, the continuous gene loss and the colinearity of most of their genomes make *Rickettsia* species an outstanding model for comparative genomics [4,6,7]. Indeed, genome reduction [8] paradoxically results in higher virulence in *R. prowazekii*.

The pathogenic mechanisms of rickettsiae are unclear. Within ticks, rickettsiae remain quiescent during the starvation of their vector but undergo a reversion to the virulent state, termed reactivation, following incubation at 37°C or blood meal [9]. This phenomenon is marked in *R. rickettsii* by morphological changes in the microcapsular and slime layers [9]. The precise molecular mechanisms of this change, however, are only poorly understood. During human infection, attachment to and invasion of host cells were suggested to involve the outer membrane proteins rOmpA and rOmpB and the adhesins Adr1 and Adr2 [10,11]. A phospholipase D activity was proposed to play a role in escape from phagosomes [8,12], and intracellular motility was demonstrated to rely on actin polymerization [13,14]. None of these factors nor the presence of a type IV secretion system [15], however, explain the virulence differences observed among *Rickettsia* species [6].

Over the last ten years, *R. africae* has emerged as the causative agent of African tick-bite fever [2], the most common SFG rickettsiosis both in terms of seroprevalence [16] and incidence [17-20]. Such an epidemiologic success is due to various factors, including the increase of tourism to wildlife parks in sub-Saharan Africa, the attack host-seeking behavior of its vector ticks, *Amblyomma* sp., and the elevated prevalence of *R. africae* in these ticks, with infection rates of up to 100% [21]. In addition, the bacterium has been identified in other areas with warm climates, such as the West Indies, where it was found in Guadeloupe, Martinique, St Kitts and Nevis, and Antigua islands [2]. Such a distribution, as well as the presence of *R. africae* in Reunion island, is likely to result from the transfer from Africa of cattle bearing infected ticks [2]. Tick-associated rickettsiae may infect ticks feeding on infected hosts or may be passed from one generation to the next transovarially. *R. africae* is transmitted transovarially and appears to be the most successful rickettsia in its adaptation to its vector tick, as the prevalence of tick infection is higher than that of any other rickettsia [22]. In addition, infection does not appear to alter tick fitness (P. Parola, unpublished data). These data highlight the fact that *R. africae* is an extremely successful and fit bacterium.

By comparison with *R. conorii*, the second most prevalent SFG rickettsia in Africa, whose genome has previously been sequenced [23], *R. africae* exhibits a higher prevalence in ticks [2], a lower virulence in humans [17], and a greater genetic homogeneity [24]. The genetic factors underlying these characteristics are, however, unknown. We assumed that the *R. africae* genome sequence might help understand the characteristics of this species and the genetic mechanisms associated with the difference in virulence. Here, we present the sequence of the *R. africae* genome and additional data that suggest that this species has emerged recently. In support of this hypothesis, we show that *R. africae* is a clonal population. We also present data that support the assumption that rickettsial virulence increases following gene inactivation.

Results

General Features of the Genome

The genome of *R. africae* consists of two replicons: a circular chromosome of 1,278,540 base pairs (bp) (Figure 1) and a 12,377 bp circular plasmid (Table 1, Figure 2[25,26]). We acknowledge the fact that the ESF-5 strain, first isolated in 1966 [27], may have undergone loss or rearrangement of plasmid or chromosomal genes during multiple passages in cell culture. Sequences were deposited in GenBank under accession number [GenBank: NZ_AAUY01000001]. The chromosome has a G + C content of 32.4%, in the range of other SFG rickettsial genomes (32.3 – 32.5%), whereas the plasmid has a G + C content of 33.4%, similar to those of *R. felis* (33.2 and 33.6%) [28] but higher than that of *R. massiliae* plasmids (31.4%). The predicted total complement of 1,271 open reading frames (ORFs), 1,260 chromosomal (78.26% coding sequence), and 11 plasmidic (81.3% coding sequence) ORFs [see Additional file 1], is in the range of genomes from SFG rickettsiae with the exception of *R. felis*, which exhibits a larger genome (Table 1). Of these, 1,117 (87.9%) exhibited homologs in the non-redundant database, and 1,024 (80.5%) were assigned putative functions [see Additional file 2]. Overall, the 1,260 chromosomal ORFs encoded 1,112 protein-coding genes, with 87 of these being split into 2 to 10 ORFs by the presence of one to several stop codons. By comparison with other SFG genomes, *R. africae* had fewer split genes than any other species with the exception of *R. felis* (Table 1). In addition, *R. africae* exhibited a single rRNA operon, with non-contiguous 16S and 23S rRNA genes as in other rickettsial genomes, 33 tRNAs and another three RNAs. The *R. africae* chromosome exhibited an almost perfect colinearity with the *R. conorii* genome [23], with the exception of a 88,459-bp inversion [see Additional file 3]. At both extremities of the inversion, there were repeats of the *Rickettsia* palindro-

Table 1: General features of *Rickettsia* genomes.

Species (strains)	Genome size (bp)	G+C content (%)	Protein-coding genes	RNAs	References
Spotted fever group					
R. africae (ESF-5)	1,290,917	32.4	1,123	39	Present study
Chromosome	1,278,540	32.4	1,112	39	
Plasmid pRA	12,377	33.4	11	0	
R. akari (Hartford)	1,23106	32.3	1,259	35	*
R. conorii (Malish 7)	1,268,755	32.4	1,374	39	[23]
R. felis (URRWXCal2)	1,587,240	32.5	1,512	39	[28]
Chromosome	1,485,148	32.5	1,444	39	
Plasmid pRF	62,829	33.6	68	0	
Plasmid pRF	39,268	33.2	44	0	
R. massiliae (Mtu5)	1,376,184	32.5	1,192	39	[29]
Chromosome	1,360,898	32.5	1,180	39	
Plasmid pRMA	15,286	31.4	12	0	
R. rickettsii (Sheila Smith)	1,257,710	32.5	1,345	36	*
R. rickettsii (Iowa)	1,268,175	32.4	1,384	37	[25]
R. sibirica (246)	1,250,021	32.5	1,083	36	*
Typhus group					
R. prowazekii (Madrid E)	1,111,523	29.0	834	39	[58]
R. typhi (Wilmington)	1,111,496	28.9	838	39	[26]
Third group					
R. bellii (RML369-C)	1,522,076	31.7	1,429	40	[15]
R. bellii (OSU 85–389)	1,528,980	31.6	1,476	36	*
R. canadensis (McKiel)	1,159,772	31.1	1,093	36	*

* Unpublished genomes available in GenBank: *R. akari* (Hartford) [GenBank: NC_009881],*R. rickettsii* (Sheila Smith) [GenBank: NC_009882], *R. sibirica* (246) [GenBank: NZ_AABW00000000], *R. bellii* (OSU 85–389) [GenBank: NC_009883], and *R. canadensis* (McKiel) [GenBank: NC_009879].

mic element – 6 (RPE-6) familly. In this inverted fragment, *R. africae* exhibited 20 ORFs and 10 RPEs that were absent from *R. conorii*. Among these 20 ORFs, a cluster of 11 consecutive ORFs had orthologs in the 3'-extremity of the Tra cluster previously identified in the *R. massiliae* genome [29]. These 11 ORFs included *traDF* (ORF0650), a transposase (ORF0651), *spoT*15 (ORF0652), a split *spoT*13 (ORF0653/ORF0654), a split *spoT*6 (ORF0655/ORF0656), a split signal transduction histidine kinase (ORF0657/ORF0658), *dam2*, a site-specific DNA adenine methylase (ORF0659), and ORF0660 of unknown function (Figure 3). In addition to the orthologs in *R. massiliae*, these genes had orthologs in similar clusters in *R. felis*, *R. bellii*, *R. canadensis* and *O. tsutsugamushi* but were absent from all other species. As in *R. massiliae*, *R. bellii* and *R. canadensis*, the *R. africae* cluster was bounded at its 3'-end by a tRNA-Val, but, in contrast with these three species, neither an integrase with its *attI* site nor a tRNA-Val fragment marker of integration was present at the 5' end (Figure 3). The presence of a similar gene cluster inserted at the same position in several *Rickettsia* species, with a GC content different from that of the genome (29.78% *vs* 32.4%, respectively, in *R. africae*) suggests that it was acquired horizontally from a common ancestor and then transmitted vertically. In *R. africae*, an *attC* site, specific to integron-inserted gene cassettes, located at the 3'-end (coordinates: 687890–688018) of the *spoT*15 gene (ORF652), supports the role of integration in the insertion of this gene cluster. *AttC* sites were also identified in *R. massiliae* (coordinates: 743029–743145), *R. felis* (coordinates: 407889–408017), and *R. bellii* (coordinates 468143–468211). Nevertheless, the presence of transposases in all species and the fact that, in *R. felis*, nine of these genes are located in the pRF plasmid support the role of several genetic mechanisms at the origin of this cluster, possibly involving plasmids, integrons and transposons. In comparison with other species containing this gene cluster, *R. africae* had the smallest number of genes. In particular, it lacked most of the Tra cluster, with the exception of *traDF*, but retained three *spoT* genes, including two degraded to pseudogenes. In *R. bellii* and *R. massiliae*, *tra* genes were described as encoding components of a type IV secretion system (T4SS) for conjugal DNA transfer [15,29]. In terms of gene content, the *R. africae* cluster was more similar to those of *R. felis* and *R. canadensis*, with the loss of the Tra cluster, the conservation of *spoT* genes and the presence of pseudogenes, than to those of *R. massiliae* and *R. bellii*, in which the Tra cluster was intact but *spoT* genes were partially degraded. Such findings suggest that species-specific evolution of this gene cluster

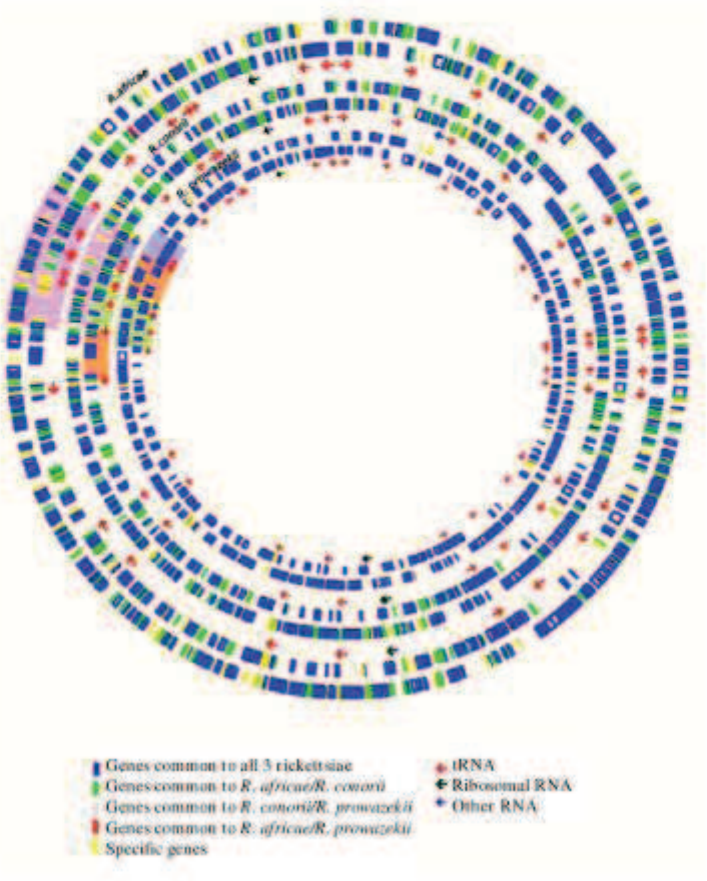


Figure 1
Circular representation of the genomes of *R. africae*, *R. conorii*, and *R. prowazekii* based on data from GenBank entries [GenBank: NZ_AAUY01000001], [GenBank: NC_003103] **and** [GenBank: NC_000963], **respectively**. Protein coding genes common to all species are in blue; genes common to *R. africae* and *R. conorii* are in green; genes common to *R. africae* and *R. prowazekii* are in red; genes common to *R. conorii* and *R. prowazekii* are in pink and specific genes in each genomes are in yellow. Common genes are identified using best BLAST match. The region of rearrangement of the genome between *R. africae* and *R. conorii* is colored in purple; the regions of rearrangment between *R. prowazekii* and *R. conorii* are colored in orange, light green, yellow and light blue. Also represented are transfer RNAs (red arrows), ribosomal RNAs (dark arrows) and other RNAs.

occurred, which likely resulted from gene excisions in *R. africae*, *R. felis* and *R. canadensis*, or gene expansion by transposase duplication in *R. massiliae*.

In addition to the $traD_I$ gene described above, the *R. africae* chromosome retained many of the components of the type IV secretion system (T4SS) involved in both DNA transfer and effector translocation in other bacteria [30], including *virB1*, *virB2* (ORF0232), *virB3* (ORF0128), *virB4* (ORF0129, ORF1109), *virB6* (ORF0130, ORF0131, ORF0132, ORF0133, ORF0134, ORF0135), *virB8* (ORF0359, ORF0361), *virB9* (ORF0358, ORF0362), *virB10* (ORF0363), *virB11*(ORF0364), and *virD4* (ORF0365). In addition, *R. africae* possessed a *traX* (ORF0816) and a split *fimD* (ORF0592/ORF0593/ ORF0594) gene but lacked other Tra cluster genes found in *R. massiliae*, *R. felis*, *R. bellii* and *O. tsutsugamushi*, such as *traC* and $traG_I$[15,28,29,31]. Therefore, the Tra cluster was mostly eliminated from the *R. africae*, and, following a "use it or lose it" scheme, this species probably did not need a *tra* gene-linked conjugation system. In addition, the pRA plasmid did not contain genes encoding proteins involved in conjugation.

Six transposase-encoding genes were identified in the chromosome, including one split into two ORFs (ORF0955/ORF0956) and one present as a remnant and two in the pRA plasmid, including one present as a fragment. This contrasts with the large expansion of trans-

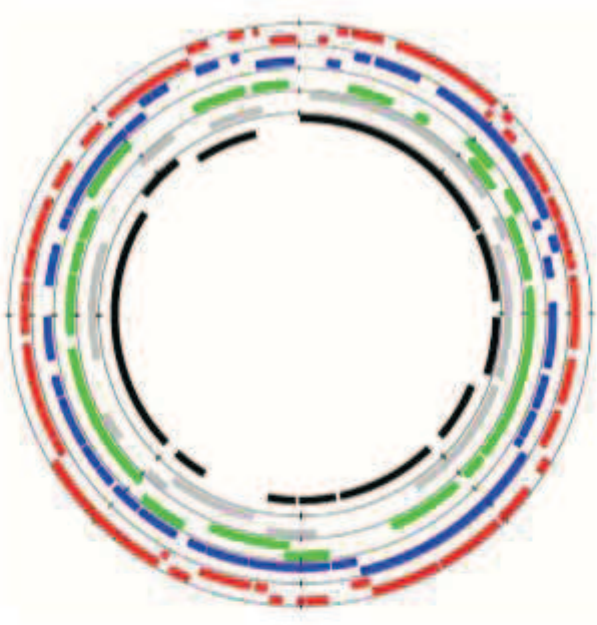


Figure 2
Circular representation of *Rickettsia* plasmids. A) The pRA plasmid: circles indicate (from the outside to the inside, on the reverse and forward strands) the GC skew, GC content, and ORFs; B) *Rickettsia* plasmids sequenced to date: circles indicate (from the outside to the inside, on the reverse and forward strands) *R. felis* pRF plasmid (red), *R. felis* pRF plasmid (blue), *R. monacensis* pRM plasmid (green), *R. massiliae* pRMa plasmid (grey), and *R. africae* pRA plasmid (black).

posases caused by gene duplications previously detected in *R. felis* and *R. bellii* [15,28].

Common rickettsial gene set and phylogeny

When compared to eight other available rickettsial genomes, a total of 645 genes and 39 RNA-encoding genes of *R. africae* had orthologs in all genomes. In addition, another 32 *R. africae* genes had orthologs only in SFG rickettsiae and were either absent or remnant in TG rickettsiae. Consequently, we identified 645 genes as constituting the core gene set of all available rickettsial genomes and 700 ORFs as the core gene set of SFG rickettsiae. Following concatenation of the 645 core genes, a reliable phylogenetic organization (Figure 4) was obtained using three analysis methods that was consistent with previous phylogenetic studies of *Rickettsia* species [4,32-36].

In comparison with other *Rickettsia* genomes, *R. africae* had 242, 238 and 69 fewer genes than *R. bellii*, *R. felis* and *R massiliae*, respectively, but 279, 260, 52, 23, 17, and 15 more genes than *R. typhi*, *R. prowazekii*, *R. akari*, *R. rickettsii*, *R. sibirica*, and *R. conorii*, respectively. When comparing the numbers of degraded genes (split + remnants), *R. africae*, with 127 degraded genes, had a significantly less degraded genome ($P < 10^{-2}$) than that of other spotted fever group rickettsiae including *R. akari* (176), *R. conorii* (196), *R. massiliae* (212), *R. rickettsii* (198) and *R. sibirica* (199) (Table 1). It had, however, significantly more degraded genes than *R. felis* (86, $P < 10^{-2}$).

Transcription of genes conserved in R. africae *but absent from highly pathogenic species*

R. africae had 18 intact genes that were either absent or degraded in all three virulent species *R. conorii*, *R. rickettsii* and *R. prowazekii*. Of these, 12 encoded proteins of unknown functions (raf_ORF0036, raf_ORF0064, raf_ORF0391, raf_ORF0412, raf_ORF0414, raf_ORF0415, raf_ORF0445, raf_ORF0660, raf_ORF0758, raf_ORF0793, raf_ORF0876, and raf_ORF0884) (Figure 5) [see Additional file 4]. The remaining six genes encoded a plasmid maintenance system antidote protein (raf_ORF0424), the *spoT15* gene (raf_ORF0652), a site-specific DNA adenine methylase (Dam2) (raf_ORF0659), an ankyrin repeat (raf_ORF0782), a putative integral membrane protein (raf_ORF0973), and a protein (RIG1002) exhibiting a high degree of amino acid sequence identity (>50%) with proteins of -proteobacteria classified within the COG3943 as putative virulence proteins. When investigating the transcription of these 18 genes in *R. africae* grown at 28, 32 and 37°C, we observed a significantly higher transcription level at 37°C than at lower temperatures for two genes, raf_ORF414 and raf_ORF660. The

Figure 3
Presence of the *tra* gene cluster in *Rickettsia* species. Raf: *R. africae*; Rma: *R. massiliae*; Rfe: *R. felis*; Rbe: *R. bellii*; Rco: *R. conorii*; Rsi: *R. sibirica*; Rri: *R. rickettsii*; Rak: *R. akari*; Rpr: *R. prowazekii*; Rty: *R. typhi*.

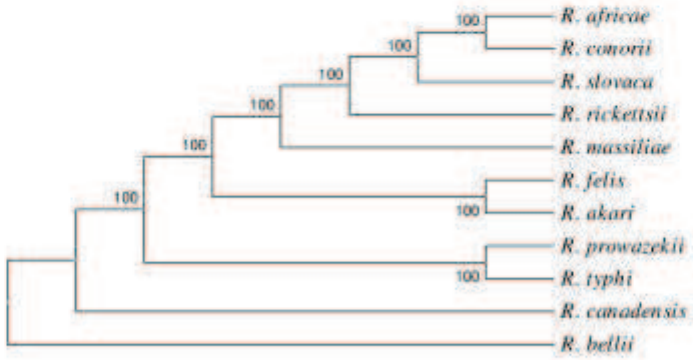
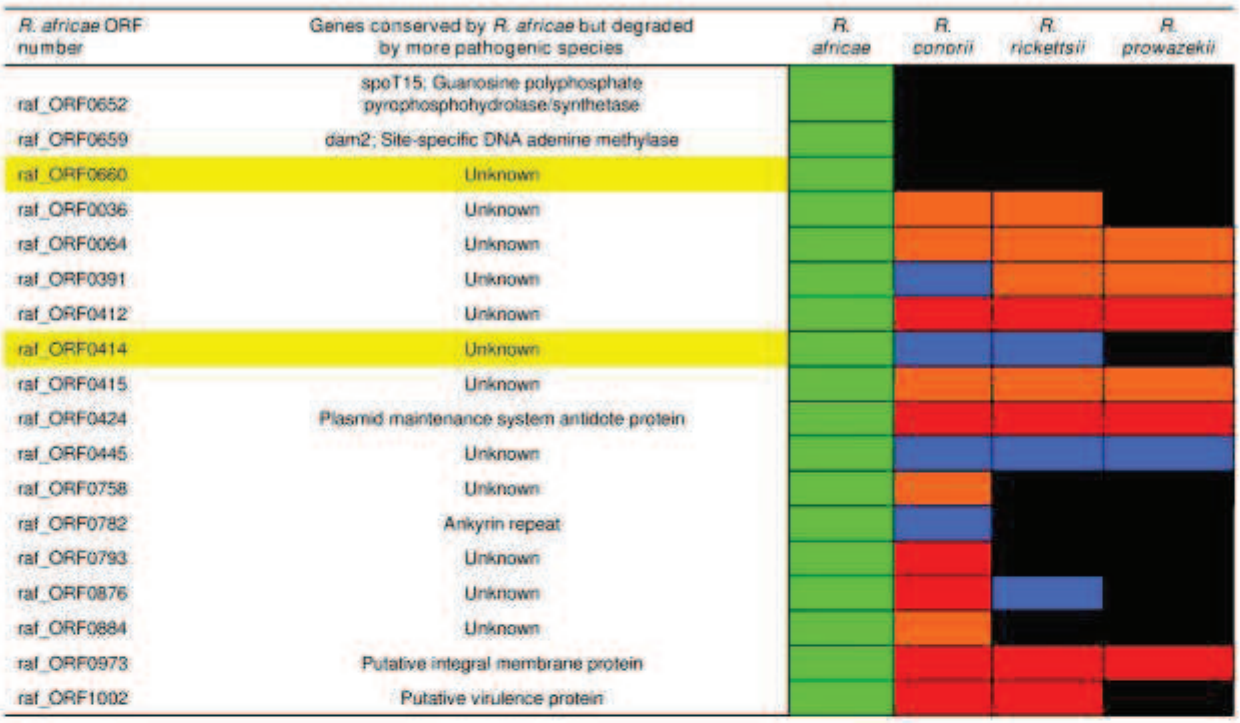


Figure 4
Phylogenetic tree inferred from the comparison of 645 concatenated *Rickettsia* core protein-coding genes. Similar organizations were obtained using both the maximum parsimony and neighbor joining methods. Bootstrap values are indicated at branch nodes.



R. africae ORF number	Genes conserved by *R. africae* but degraded by more pathogenic species
raf_ORF0652	spoT15; Guanosine polyphosphate pyrophosphohydrolase/synthetase
raf_ORF0659	dam2; Site-specific DNA adenine methylase
raf_ORF0660	Unknown
raf_ORF0036	Unknown
raf_ORF0064	Unknown
raf_ORF0391	Unknown
raf_ORF0412	Unknown
raf_ORF0414	Unknown
raf_ORF0415	Unknown
raf_ORF0424	Plasmid maintenance system antidote protein
raf_ORF0445	Unknown
raf_ORF0758	Unknown
raf_ORF0782	Ankyrin repeat
raf_ORF0793	Unknown
raf_ORF0876	Unknown
raf_ORF0884	Unknown
raf_ORF0973	Putative integral membrane protein
raf_ORF1002	Putative virulence protein

Figure 5
Schematic representation of the genes conserved in *R. africae* but lost by highly pathogenic rickettsiae. Genes highlighted in yellow are upregulated at 37°C. The state of a gene is represented by a small box colored in green (full-length), blue (pseudogene), red (fragment), orange (remnant) or black (absent). Gene numbers are indicated in the left column.

former gene contained a putative protease domain site, but the latter had no known function.

The R. africae *plasmid*

The *R. africae* plasmid (Figure 2) is a new example of a plasmid in *Rickettsia* species, following those in *R. felis* [28], *R. massiliae* [29], *R. monacensis* [37], *R. helvetica*, *R. peacockii*, *R. amblyommii* and *R. hoogstraalii* [38]. This plasmid, named pRA, is smaller (12,377 bp) than those of *R. felis* (62,829 bp and 39,263 bp long, for pRF and pRF , respectively), *R. monacensis* (23,486 bp), and *R. massiliae* (15,286 bp). The pRA plasmid is predicted to contain 11 genes, 6 of which (54%) have homologs in public databases and are associated with functional attributes. These six genes encode for a chromosomal replication initiator DnaA-like protein (ORF1260), a site-specific recombinase (ORF1262), two contiguous transposases exhibiting 100% sequence similarity (ORF1263 and 1264) but with one (ORF1263) shorter than the other, the auto-transporter protein SCA12 (ORF1268), and a ParA-like plasmid stability protein (ORF1270). Five genes (ORFs 1260, 1263, 1264, 1269 and 1270) have orthologs in the *R. massiliae* plasmid, six have orthologs in the *R. felis* plasmids (ORF1260, 1263, 1264, 1268, 1269 and 1270), and three have orthologs in the *R. monacensis* plasmid (ORF1260, ORF1268, and ORF1270). The presence of two genes (ORF1260 and 1270) conserved in plasmids from four species suggests that these plasmids have a common origin. The presence of two almost identical successive transposases in *R. africae* matching a single gene in *R. massiliae* and *R. felis* suggests a duplication event in the former species. The pRA plasmid lacks heat shock protein-encoding genes found in other rickettsial plasmids. In contrast, ORF1262, a site-specific recombinase, is absent from other species. Its closest phylogenetic neighbour is a site-specific recombinase from *Magnetospirillum magnetotacticum*, a high G-C content -proteobacterium living in aquatic environments [39]. The *sca12* gene (ORF) found intact in *R. africae* pRA was absent from the *R. massiliae* and *R. monacensis* plasmids and present but fragmented within *R. felis* pRF, but it was absent from pRF as well all other *Rickettsia* species.

As outlined by Baldridge *et al.* [38], the plasmid content of a *Rickettsia* species may vary according to the passage history of rickettsial strains. When estimating the prevalence

of the plasmid among *R. africae* strains, we detected it in the 22 tested isolates from South Africa and in the 48 eschar biopsies from patients with ATBF contracted in the same country and in 20/32 *R. africae*-positive *Amblyomma* ticks [see Additional files 5 and 6]. Therefore, it appears from these results that, depending on the geographic location, the plasmid of *R. africae* may be unstable. Whether the plasmid has been lost by PCR-negative strains or cannot be amplified with the primers we used is as yet unknown. Such inter-strain differences in plasmid content were also observed in *R. felis* (Unpublished data).

Stress response

Rickettsiae live intracellularly in both arthropod and mammal hosts. This implies that periods of tick starvation and feeding cause bacterial dormancy and multiplication following reactivation [40]. As a consequence, and despite their obligate intracellular location, rickettsiae may face, and thus have to adapt to, highly variable and extreme environmental conditions. Known as the stringent response, this bacterial adaptation to nutritional stress has been described to be mediated by the accumulation of guanosine nucleotides pppGpp (guanosine 3'-diphosphate 5'-triphosphate) and ppGpp (guanosine 3'-diphosphate 5'-diphosphate) [41]. Accordingly, the transcriptional analysis of *R. conorii* exposed to a nutrient deprivation was characterized by the up-regulation of *gmk* and of genes from the *spoT* family, suggesting a role for these nucleotides as effectors of the stringent response [42,43]. The *R. africae* genome exhibited eight *spoT* genes phylogenetically classified within two major clades [see Additional file 7]. The largest clade included *spoT* genes with hydrolase activity (1–10, 14, 15, 17–21), while the second included those with a synthetase domain. With eight genes, *R. africae* had more *spoT* genes than *R. rickettsii* (5 genes), *R. conorii* (4), *R. sibirica* (4), *R. akari* (7), *R. canadensis* (5), *R. typhi* (4) and *R. prowazekii* (1) but fewer genes than *R. felis* (14) and *R. bellii* (10) [see Additional file 8]. Altogether, our data suggest that *R. africae* is more regulated than more pathogenic species.

Infection of mammal hosts

The *R. africae* genome encoded rOmpA (or Sca0) and rOmpB (or Sca5), two surface-exposed and immunodominant proteins belonging to the paralogous "surface cell antigen" (SCA) family and known in *Rickettsia* species to be responsible for antigenic differences between species [1] and to elicit an immune response in patients [44]. Experimental studies suggested that these two auto-transporter proteins could function as adhesins [10,11,45,46]. In addition, another eight SCA-encoding genes were found in the genome. These 10 genes were represented by 22 ORFs due to partial degradation of some of the paralogs [see Additional file 8]. Among the 17 SCA-encoding genes detected in *Rickettsia* species [47], *R. africae* had similar sets of conserved (*sca*0 – 2, 4 and 5), degraded (*sca*3, 8 – 10 and 13) and absent (*sca*6, 7, 11, 14 – 17) *sca* genes as *R. conorii* and *R. rickettsii*. In addition to these 10 SCA-encoding genes, *R. africae* exhibited a degraded *sca*9 gene and a complete *sca*12 gene carried by the pRA plasmid, only shared with *R. felis*, where it was also found partially degraded on the pRF plasmid. The *sca*12 genes from both species were grouped into a distinct cluster close to the *sca*1, 2 and 6 genes [see Additional file 9]. This result further supports a common origin of the pRA and pRF plasmids.

A proteomic approach recently allowed the identification of two paralogous proteins encoded by the genes RC1281-RC1282 and RP827-RP828, as putative adhesins Adr1 and Adr2. These proteins may be key actors for entry and infection in both *R. conorii* and *R. prowazekii* [11]. Both proteins are ubiquitously present within the *Rickettsia* genus [4]. Their presence within the *R. africae* genome (ORF1174 + ORF1175) [see Additional file 10] reinforces their suspected key role in rickettsial life.

Both *pld* and *tlyC*, encoding phospholipase D [8] and hemolysin C [12], respectively, which play a role in phagosomal escape [13,48], were conserved in the *R. africae* genome (ORF1161 and ORF1039, respectively). This bacterium also exhibited genes encoding other proteins with membranolytic activity, including *tlyA* (hemolysin A) and *pat*1 (patatin-like phospholipase) [12,49]. As expected, the genome of *R. africae* has a *rickA* gene (ORF0824) orthologous to all rickettsial *rickA* genes and coding a protein activating the Arp2/3 complex, whose nucleation triggers actin polymerisation [50] [see Additional file 11]. The Rick A protein in *R. africae* is slightly different from those of other species, with a phenylalanine instead of a serine within the G-actin-binding site, an ENNIP [PS] motif repeated twice instead of four times in the central proline-rich region of the protein [see Additional file 11], and an aspartate and an isoleucine instead of an asparagine and an alanine or valine, respectively, in the carboxy-terminal region. Despite these differences, the RickA protein of *R. africae* appeared to be functional as demonstrated by its ability to polymerize actin and multiply intranuclearly (Figure 6).

Sixteen *vir* gene paralogs were found in the *R. africae* genome. Virulence genes of the *vir* family belong to the type IV secretion machinery, a system that allows the delivery of virulence factors from bacterial and eukaryotic host membranes to the cytoplasm of the host cell [51]. All 16 genes were found to be intact and common to all *Rickettsia* genomes with the exception of *virB6-2* in *R. africae* and *virB6-5* in *R. massiliae* [see Additional file 8]. In both species, these genes were split into two ORFs. Phylogenetic analysis of the *virB6-2* gene distinguished clearly the

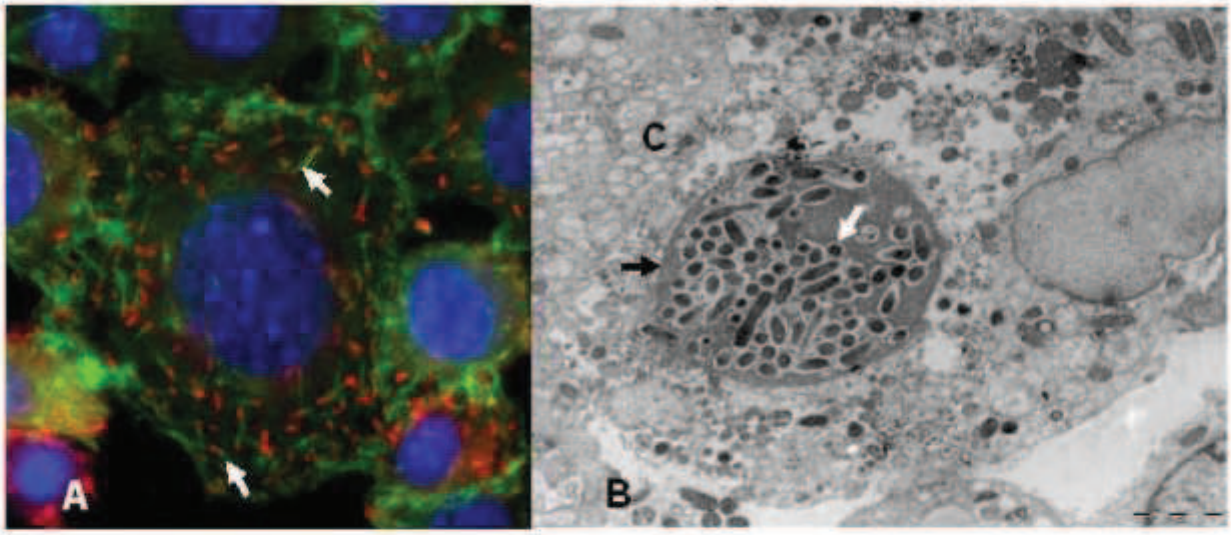


Figure 6
Intracellular motility of *R. africae*. A) Actin tail formation by *R. africae*. L-929 cells were infected with *R. africae*, fixed and stained with fluorescent phalloidin (green) and a polyclonal antibody against *R. africae* and visualized using anti-rabbit-Alexa549 as a secondary antibody (red). The white arrows show actin tails. B)*R. africae* in the cytoplasm and nucleus of L-929 cells. C = cytoplasm; black arrow = nucleus; white arrows = *R. africae* bacilli. Transmission electron microscopy. Scale bar = 5 m.

SFG and TG and showed that the *R. africae* VirB6-2 protein is phylogenetically closer to that of *R. sibirica* [see Additional file 12].

Clonality of R. africae

Of the 155 *Amblyomma* ticks tested, 139 (89.6%) were PCR-positive for *R. africae* [see Additional file 5]. Therefore, infection rates of *Amblyomma* ticks with *R. africae* may be higher than previously described [21,22,52], which suggests an extreme fitness of this rickettsia for its vector. In addition, such infection rates are the highest among *Rickettsia* species [see Additional file 13].

Using MST, PCR products of the expected sizes were obtained from the *dksA-xerC*, *mppA-purC* and *rpmE*-tRNA^{fMet} intergenic spacers from all tested specimens. Sequences obtained from these amplicons were in all cases identical to those previously obtained for *R. africae* [GenBank: DQ008280], [GenBank: DQ008301], and [GenBank: DQ008246], for the *dksA-xerC*, *mppA-purC* and *rpmE*-tRNA^{fMet} spacers, respectively). This is the first rickettsia demonstrated to be clonal. Other tested *Rickettsia* species, including *R. conorii* (31 MST genotypes out of 39 strains tested [53]), *R. massiliae* (2/7 [24]), *R. sibirica* (3/3 [24]), and *R. felis* (3/6 [24]), were significantly more genetically variable than *R. africae* ($p < 10^{-2}$ in all cases).

Discussion

Using a comparative study of rickettsial genomes, we found that virulence in *Rickettsia* species is not correlated with acquisition of foreign DNA but may rather result from a reduction in regulation due to genome decay [6,23]. Comparative genomics sheds light on a much wider spectrum of virulence acquisition mechanisms in bacteria than initially thought [54]. Based on the examples of enterobacteria and staphylococci, gain in pathogenicity in bacteria was mainly thought to result from horizontal gene transfer, either directly or through mobile genetic elements [55,56]. However, a recent study of *Rickettsia* species associated with arthropods, insects, leeches and protists clearly demonstrated that horizontal gene transfer was a rare event within this genus [5]. In addition, genomic studies demonstrated that rickettsiae are undergoing genome decay, affecting in priority horizontally-acquired genes [57], and that there is no association between pathogenicity and acquisition of virulence markers [6]. In fact, the genome of the most virulent species, *R. prowazekii* [58], is a subset of the less pathogenic species *R. conorii* [23], thus highlighting a paradoxical relationship between smaller genome size and higher pathogenicity. Careful comparison of the *R. prowazekii* and *R. typhi* genomes also demonstrated that the former species, more pathogenic than the latter, had a more decayed genome despite a 12-kb insertion that likely resulted from a single genetic event [59].

When investigating the genomic characteristics associated with the milder virulence of *R. africae*, we first ruled out a potential role of the plasmid by the fact that it is unstable in this species. Then, we compared the gene contents of *R. africae* with *R. conorii*, *R. rickettsii*, and *R. prowazekii*, which exhibit a higher pathogenicity in humans and their arthropod hosts. We observed that *R. africae* showed no gene loss but had 18 genes fully conserved that were either absent or degraded in the other species (Figure 5). We speculated that, because *R. africae* had more intact genes than more virulent species, some of these genes may be involved in maintaining a low virulence level. Such a behavior may not be unique to rickettsiae. It was found that gene knockout resulted in increased virulence in *Mycoplasma*, *Streptococcus pyogenes*, and *Vibrio cholerae* [60-62]. In *M. ulcerans*, genome reduction was also linked to gain in virulence [63]. It emerges as a concept that virulence may be increased by gene loss [54]. We assume that a similar phenomenon may happen in rickettsiae, and that inactivation of some genes may deregulate the control of bacterial multiplication, in particular during the reactivation phenomenon following warming, thus enhancing pathogenesis.

Among the 18 putative candidate genes unique to *R. africae*, we identified only two genes (raf_ORF414 and raf_ORF660) that were significantly more transcribed at 37°C than at lower temperatures. Of these, one (raf_ORF414) encoded a protein that had a putative protease domain. A protease was previously shown in *Vibrio cholerae* to be a virulence repressor [60]. However, whether this differentially-transcribed protease plays a role in virulence repression in *R. africae* is as yet unknown. In contrast, the *spoT*15 gene (raf_ORF652) unique to *R. africae* was not upregulated, and this species retained another two *spoT* pseudogenes (raf_ORF653–654 and raf_ORF655–656) that were completely lost by other species. *SpoT* genes, effectors of the stringent response, were shown to play a major role in adaptation to stress in *R. conorii*, in particular when subjected to abrupt temperature variations similar to those occurring during a tick blood meal [42]. *R. africae*, however, has more *spoT* genes than *R. conorii* or *R. rickettsii* and does not show any modification of expression of its specific *spoT*15 gene during temperature variations. We speculate that higher regulation ability in *R. africae* is linked to lower pathogenicity.

In addition, when compared to other tick-borne *Rickettsia* species, *R. africae* exhibited several unique characteristics. First, this species is extremely successful and fit: it is highly adapted and harmless to its tick host, being efficiently transmitted both transtadially and transovarially in *Amblyomma* sp. ticks, which consequently act as efficient reservoirs [64]. In contrast, *R. rickettsii* [65,66] and *R. conorii* [67] have a negative effect on their tick vectors in experimental models. As a result, the prevalence of *R. africae* in its host ticks is higher than that of most other rickettsiae. Similarly, *R. africae* is less pathogenic for humans than other SFG species such as *R. conorii* and *R. rickettsii*, in particular because the infection is never lethal [17]. This observation was later supported by the demonstration that inoculation eschars in ATBF were histologically different from those in MSF [68]. In particular, in contrast with other SFG rickettsioses where eschars are characterized by perivascular infiltration of T cells and macrophages, with some B lymphocytes and few polymorphonuclears, the vasculitis in ATBF is made of a large infiltrate of neutrophils causing an extensive cutaneous inflammation and necrosis [see Additional file 14] [68]. Such a local reaction, in addition to the few *R. africae* cells detected in eschars [68], suggests that the bacterium replicated poorly in human tissues. Second, *R. africae* has significantly fewer degraded genes than other SFG species ($p < 10^{-2}$), except *R. felis*. Specifically, this characteristic suggests that *R. africae* is undergoing a slower degradation process than other rickettsiae. Third, the identification of a single MST genotype among 102 strains suggested that *R. africae* was clonal [24,69]. This contrasted with the variable plasmid content of this species. Originally thought to be absent in *Rickettsia* species, plasmids have been detected in eight species to date [28,29,37,38], and their plasmid content may exhibit intraspecies variability. In *R. felis*, two plasmid forms have been sequenced [28], and Baldridge *et al.* found two plasmids in both *R. peacockii* and *R. amblyommii* [38]. In addition, these authors showed that *R. peacockii* lost its plasmids during long-term serial passages in cell culture [38]. In *R. africae*, the pRA plasmid may also be unstable, as shown by the absence of plasmid detection in 12/32 *Amblyomma* ticks tested. This plasmid encodes 11 ORFs, two of which are common to *R. felis*, *R. massiliae* and *R. monacensis* plasmids [see Additional file 1], which strongly suggests a common source for these mobile elements. We suspect that rickettsial plasmids and Tra clusters are vertically inherited but are apparently unstable and are currently degrading.

Conclusion

Based on its genome and lifestyle, we suspect that the clonal *R. africae* is more regulated and more specifically adapted to its host and warm environment than other tick-associated rickettsiae. We speculate that losing this regulation, as observed in several intracellular pathogens, is a critical cause of virulence [6]. Further transcriptomic analysis of *R. africae* and other *Rickettsia* species grown at various temperatures is currently ongoing to identify putative other candidate genes involved in stress response.

Methods

Genome Sequencing

Bacterial purification and DNA extraction

In this study, we used *R. africae* ESF-5 strain, CSUR R15 (Collection de souches de l'Unité des Rickettsies, Marseille, France), which was isolated in an *Amblyomma variegatum* tick collected from cattle in the Shulu province of Ethiopia in 1966 [27]. *R. africae* was cultivated in Vero cells growing in MEM with 4% fetal bovine serum supplemented with 5 mM L-glutamine. Bacterial purification, DNA extraction and pulsed-filed gel electrophoresis were performed as described in Additional file 15 [see Additional file 15].

Shotgun sequencing of R. africae genome

Three shotgun genomic libraries were made by mechanical shearing of the DNA using a Hydroshear device (GeneMachine, San Carlos, CA, USA). Sequence blunt ends, to which the BstXI adaptator was linked, were obtained using the T4 DNA polymerase (New England Biolabs). Fragments of 3, 5, and 10 kb were separated on a preparative agarose gel (FMC, Rockland, ME, USA), extracted using the Qiaquick kit (Qiagen, Hilden, Germany), and ligated into a high copy-number vector pCDNA2.1 (Invitrogen, Carlsbad, CA, USA) for the two smaller inserts and into the low copy-number vector pCNS [28] for the largest inserts. Further details are available in Additional file 15 [see Additional file 15].

Annotation

We predicted protein-coding genes (ORFs) using SelfID as previously described [15]. Functional assignments for the ORFs were based on database searches using BLAST [70] against UniProt [71], NCBI/CDD [72], and SMART [73] databases. In most cases, we applied an E-value threshold of 0.001 for the database searches to retrieve homologues. Detailed analyses using multiple sequence alignments and phylogenetic reconstructions were carried out to assign putative functions to the ORFs, when needed. Orthologous gene relationships between *R. africae* and other *Rickettsia* species were approximated using the best reciprocal BLAST match criterion. The numbers of transposases, ankyrin/tetratricopeptide repeat-containing genes, and integrases were computed using RPS-BLAST with NCBI/CDD entries related to those domains with a 10^{-5} E-value threshold. tRNA genes were identified using tRNAscan-SE [74]. To identify *Rickettsia* palindromic elements, we used hidden Markov models [75] based on the previously identified *Rickettsia* palindromic element sequences. ClustalW [76], T-coffee [77], and MUSCLE [78] were used to construct multiple sequence alignments. Toxin-antitoxin genes were identified using the Rasta-Bacteria software http://genoweb.univ-rennes1.fr/duals/RASTA-Bacteria.

Phylogenetic analysis

We based our analysis on the 645 complete orthologous genes found by Blast programmes in all *Rickettsia* genomes [70]. Subsequently, the amino acid sequences of these 645 proteins were concatenated for each genome and multiple alignment was performed using the Mafft software [79]. Gapped positions were removed. The maximum parsimony and neighbor joining trees were constructed using the MEGA 3.1 software [80].

Clonal origin of R. africae

We examined *R. africae* within 155 *Amblyomma* sp. ticks and eggs from various geographical origins [see Additional file 5]. These included 80 adults (40 male and 40 female), 40 larvae, 15 nymphs and 20 eggs. PCR amplification of the *traD* gene was performed using the *R. africae*-specific primer pair traD-F (5'-caatgcttgatctatttggtag-3') and traD-R (5'-cttccttttctctaagctatgc-3') and the probe traD-probe (5'-FAM-ttatggtgctaactccatgcgtgatg-TAMRA-3'). The presence of the plasmid was estimated using the primer pair 1267F (5'-ccagccattaccgtaatcac-3') and 1267R (5'-tagtgccttatactcaagttc-3') and the probe 1267-probe (5'-FAM-gcagaaagtgattaaggcgatcagctg-TAMRA-3') that is able to detect ORF 1267 encoding a protein of unknown function specific to the plasmid. The presence of the plasmid was examined in 22 strains obtained from patients who contracted the disease in South Africa and maintained in the CSUR [see Additional file 6], in PCR-positive eschar biopsies from another 48 patients who developed ATBF following a trip to South Africa, and in 32 *Amblyomma* sp. ticks found positive for *R. africae*, using the above-described PCR assay [see Additional file 5]. To evaluate the genetic diversity of *R. africae*, we used the multi-spacer typing (MST) method as previously described [53]. This method has been described as the most discriminatory genotyping tool at the intraspecies level in *Rickettsia* sp. [53]. We applied this method to the aforementioned 22 human *R. africae* strains, 48 eschar biopsies, and 32 *Amblyomma* sp. ticks from Sudan (3), Madagascar (3), Mali (3), Niger (6), Central African Republic (6), Ivory Coast (3), Guadeloupe (4), Martinique (2), and St Kitts and Nevis (2) [see Additional file 5]. The obtained sequences were compared to those available in GenBank, and the MST genotypes were determined as previously described [53].

Transcription of genes conserved in R. africae but absent from highly pathogenic species

To evaluate the transcription of the 18 genes conserved by *R. africae* and degraded in highly pathogenic species, we designed specific primer pairs and probes for each gene and tested the transcription of these genes by RT-PCR on RNA extracted from *R. africae*- infected Vero cells cultivated at 32 and then at 37°C and in XTC cells at 28 and

32°C. Experimental protocols are detailed in Additional file 15 [see Additional file 15].

Authors' contributions

PEF and DR designed the study, drafted the manuscript, and gave final approval of the submitted version; KE, QL, CR, BG, PR, CR, PP, and SA performed experiments, drafted the manuscript and gave final approval of the submitted version.

Additional material

Additional file 1

Gene content of the* R. africae *plasmid. *GenBank accession number sare indicated in square brackets. The Table includes a comparison of rickettsial plasmid contents.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-10-166-S1.doc]

Additional file 2

R. africae *gene content.* *The Table includes the gene content of the* R. africae *genome.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-10-166-S2.doc]

Additional file 3

Inversion observed by alignment of the* R. africae *(up) and* R. conorii *(down) genomes. *The Figure shows an alignment of the* R. conorii *and* R. africae *genomes.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-10-166-S3.ppt]

Additional file 4

Schematic representation of the genes diversely conserved in* R. africae *in comparison with highly pathogenic rickettsiae. *The state of a gene is represented by a small box colored in green (full-length), blue (pseudogene), red (fragment), orange (remnant) or black (absent).Gene numbers are indicated in the left column. The Figure shows the gene distribution in* R. africae *by comparison with highly pathogenic rickettsiae.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-10-166-S4.ppt]

Additional file 5

PCR-detection of* R. africae *and in* Amblyomma *ticks. *Results are indicated as number of ticks positive/number tested. The Table includes the results from PCR detection of the* R. africae *chromosome and plasmid in ticks.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-10-166-S5.doc]

Additional file 6

Rickettsia africae *strains used in this study.* *The table lists all* R. africae *strains used in this study.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-10-166-S6.doc]

Additional file 7

Phylogenetic tree showing the organization of* spoT *genes in* Rickettsia *species. *Phylogenetic relationships were inferred from aligned sequences using the Mega3.1 software with the Neighbor-Joining method. Bootstrap values are indicated at the nodes. The Figure is a phylogenetic tree showing the organization of* spoT *genes in* Rickettsia *species.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-10-166-S7.ppt]

Additional file 8

R. africae *ORFs compared to other available* Rickettsia *genomes.* *The table details the distribution of* R. africae *ORF in other rickettsial genomes.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-10-166-S8.doc]

Additional file 9

Phylogenetic tree showing the organization of* sca *genes in* Rickettsia *species. *Phylogenetic relationships were inferred from aligned sequences using the Mega3.1 software with the Neighbor-Joining method. Bootstrap values are indicated at the nodes. The Figure is a phylogenetic tree showing the organization of* sca *genes in* Rickettsia *species.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-10-166-S9.ppt]

Additional file 10

Phylogenetic tree showing the organization of* adr *genes in* Rickettsia *species. *Phylogenetic relationships were inferred from aligned sequences using the Mega3.1 software with the Neighbor-Joining method. Bootstrap values are indicated at the nodes. The Figure is a phylogenetic tree showing the organization of* adr *genes in* Rickettsia *species.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-10-166-S10.ppt]

Additional file 11

Features of RickA repeat proline-rich motif in* R. africae *and other SFG rickettsiae. *The motif " [EDGKQG]- [NS]-N- [IV]-[PSLTR]{0,28}" was used to extract these repeats using a PatternMatchingtool. The table details RickA repeat proline-rich motifs in* R. africae *and other SFG rickettsiae.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-10-166-S11.doc]

Additional file 12

Phylogenetic tree showing the organization of* virB6-2 *genes in* Rickettsia *species. *Phylogenetic relationships were inferred from aligned sequences using the Mega3.1 software with the Neighbor-Joining method. Bootstrap values are indicated at the nodes. The Figure is a phylogenetic tree showing the organization of* virB6-2 *genes in* Rickettsia *species.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-10-166-S12.ppt]

Additional file 13

Comparison of epidemiological and clinical characteristics of* Rickettsia *species. *The table includes data about the epidemiological and clinical characteristics of* Rickettsia *species.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-10-166-S13.doc]

Additional file 14

Immunohistochemical detection of* R. africae *(arrows) in the inoculation eschar of a patient with ATBF (monoclonal rabbit anti-*R. africae *antibody used at a dilution of 1:1,000 and hematoxylin counterstain; original magnification ×250). *The Figure shows the presence of* R. africae *in the inoculation eschar of a patient with ATBF, revealed by immunohistochemistry.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-10-166-S14.ppt]

Additional file 15

Supplementary material and methods. *The data provided include detailed material and methods that were used for the genome sequencing and sequence analysis of* R. africae.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-10-166-S15.doc]

Acknowledgements

This work was funded by the Network of Excellence "EuroPathoGenomics".

References

1. Raoult D, Roux V: **Rickettsioses as paradigms of new or emerging infectious diseases.** *Clin Microbiol Rev* 1997, **10**:694-719.
2. Parola P, Paddock CD, Raoult D: **Tick-borne rickettsioses around the world: emerging diseases challenging old concepts.** *Clin Microbiol Rev* 2005, **18**:719-756.
3. Perlman SJ, Hunter MS, Zchori-Fein E: **The emerging diversity of Rickettsia.** *Proc Biol Sci* 2006, **273**:2097-2106.
4. Blanc G, Ogata H, Robert C, Audic S, Suhre K, Vestris G, *et al.*: **Reductive genome evolution from the mother of *Rickettsia*.** *PLOS Genet* 2007, **3**:e14.
5. Weinert LA, Werren JH, Aebi A, Stone GN, Jiggins FM: **Evolution and diversity of Rickettsia bacteria.** *BMC Biol* 2009, **7**:6.
6. Darby AC, Cho NH, Fuxelius HH, Westberg J, Andersson SG: **Intracellular pathogens go extreme: genome evolution in the Rickettsiales.** *Trends Genet* 2007, **23**:511-520.
7. Sallstrom B, Andersson SG: **Genome reduction in the alpha-Proteobacteria.** *Curr Opin Microbiol* 2005, **8**:579-585.
8. Renesto P, Dehoux P, Gouin E, Touqui L, Cossart P, Raoult D: **Identification and characterization of a phospholipase D-superfamily gene in rickettsiae.** *J Infect Dis* 2003, **188**:1276-1283.
9. Hayes SF, Burgdorfer W: **Reactivation of *Rickettsia rickettsii* in *Dermacentor andersoni* ticks: an ultrastructural analysis.** *Infect Immun* 1982, **37**:779-785.
10. Uchiyama T: **Adherence to and invasion of Vero cells by recombinant Escherichia coli expressing the outer membrane protein rOmpB of Rickettsia japonica.** *Ann N Y Acad Sci* 2003, **990**:585-590.
11. Renesto P, Samson L, Ogata H, Azza S, Fourquet P, Gorvel JP, *et al.*: **Identification of two putative rickettsial adhesins by proteomic analysis.** *Res Microbiol* 2006, **157**:605-612.
12. Whitworth T, Popov VL, Yu XJ, Walker DH, Bouyer DH: **Expression of the Rickettsia prowazekii pld or tlyC gene in Salmonella enterica serovar Typhimurium mediates phagosomal escape.** *Infect immun* 2005, **73**:6668-6673.
13. Teysseire N, Boudier JA, Raoult D: ***Rickettsiaconorii* entry into vero cells.** *Infect Immun* 1995, **63**:366-374.
14. Gouin E, Gantelet H, Egile C, Lasa I, Ohayon H, Villiers V, *et al.*: **A comparative study of the actin-based motilities of the pathogenic bacteria Listeria monocytogenes, Shigella flexneri and Rickettsia conorii.** *J Cell Sci* 1999, **112**:1697-1708.
15. Ogata H, La Scola B, Audic S, Renesto P, Blanc G, Robert C, *et al.*: **Genome sequence of *Rickettsia bellii* illuminates the role of amoebae in gene exchanges between intracellular pathogens.** *PLoS Genet* 2006, **2**:e76.
16. Tissot-Dupont H, Brouqui P, Faugere B, Raoult D: **Prevalence of antibodies to *Coxiella burnetii*, *Rickettsia conorii*, and *Rickettsia typhi* in seven African countries.** *Clin Infect Dis* 1995, **21**:1126-1133.
17. Raoult D, Fournier PE, Fenollar F, Jensenius M, Prioe T, de Pina JJ, *et al.*: ***Rickettsia africae*, a tick-borne pathogen in travelers to sub-Saharan Africa.** *N Engl J Med* 2001, **344**:1504-1510.
18. Jensenius M, Fournier PE, Kelly P, Myrvang B, Raoult D: **African tick bite fever.** *Lancet Infect Dis* 2003, **3**:557-564.
19. Jensenius M, Fournier PE, Vene S, Hoel T, Hasle G, Henriksen AZ, *et al.*: **African tick bite fever in travelers to rural sub-Equatorial Africa.** *Clin Infect Dis* 2003, **36**:1411-1417.
20. Jensenius M, Hoel T, Raoult D, Fournier PE, Kjelshus H, Bruu AL, *et al.*: **Seroepidemiology of *Rickettsia africae* infection in Norwegian travellers to rural Africa.** *Scand J Infect Dis* 2002, **34**:93-96.
21. Parola P, Inokuma H, Camicas JL, Brouqui P, Raoult D: **Detection and identification of spotted fever group rickettsiae and ehrlichiae in African ticks.** *Emerg Infect Dis* 2001, **7**:1014-1017.
22. Tissot-Dupont H, Cornet JP, Raoult D: **Identification of rickettsiae from ticks collected in the Central African Republic using the polymerase chain reaction.** *Am J Trop Med Hyg* 1994, **50**:373-380.
23. Ogata H, Audic S, Renesto-Audiffren P, Fournier PE, Barbe V, Samson D, *et al.*: **Mechanisms of evolution in *Rickettsia conorii* and *R. prowazekii*.** *Science* 2001, **293**:2093-2098.
24. Fournier PE, Raoult D: **Identification of rickettsial isolates at the species level using multi-spacer typing.** *BMC Microbiol* 2007, **7**:72.
25. Ellison DW, Clark TR, Sturdevant DE, Virtaneva K, Porcella SF, Hackstadt T: **Genomic comparison of virulent *Rickettsia rickettsii* Sheila Smith and avirulent *Rickettsia rickettsii* Iowa.** *Infect immun* 2008, **76**:542-550.
26. McLeod MP, Qin X, Karpathy SE, Gioia J, Highlander SK, Fox GE, *et al.*: **Complete genome sequence of Rickettsia typhi and comparison with sequences of other rickettsiae.** *J Bacteriol* 2004, **186**:5842-5855.
27. Burgdorfer W, Ormsbee RA, Schmidt ML, Hoogstraal H: **A search for the epidemic typhus agent in Ethiopian ticks.** *Bull WHO* 1973, **48**:563-569.
28. Ogata H, Renesto P, Audic S, Robert C, Blanc G, Fournier PE, *et al.*: **The genome sequence of *Rickettsia felis* identifies the first putative conjugative plasmid in an obligate intracellular parasite.** *PLoS Biol* 2005, **3**:e248.
29. Blanc G, Ogata H, Robert C, Audic S, Claverie JM, Raoult D: **Lateral gene transfer between obligate intracellular bacteria: evidence from the Rickettsia massiliae genome.** *Genome Res* 2007, **17**:1657-1664.
30. Cascales E, Christie PJ: **The versatile bacterial type IV secretion systems.** *Nat Rev Microbiol* 2003, **1**:137-149.
31. Cho NH, Kim HR, Lee JH, Kim SY, Kim J, Cha S, *et al.*: **The Orientia tsutsugamushi genome reveals massive proliferation of con-**

jugative type IV secretion system and host-cell interaction genes. *Proc Natl Acad Sci USA* 2007, **104**:7981-7986.
32. Roux V, Raoult D: **Phylogenetic analysis of the genus *Rickettsia* by 16S rDNA sequencing.** *Res Microbiol* 1995, **146**:385-396.
33. Fournier PE, Roux V, Raoult D: **Phylogenetic analysis of spotted fever group rickettsiae by study of the outer surface protein rOmpA.** *Int J Syst Bacteriol* 1998, **48**:839-849.
34. Roux V, Raoult D: **Phylogenetic analysis of members of the genus *Rickettsia* using the gene encoding the outer-membrane protein rOmpB (ompB).** *Int J Syst Evol Microbiol* 2000, **50**:1449-1455.
35. Gillespie JJ, Williams K, Shukla M, Snyder EE, Nordberg EK, Ceraul SM, *et al.*: ***Rickettsia* phylogenomics: unwinding the intricacies of obligate intracellular life.** *PLOS ONE* 2008, **3**:e2018.
36. Williams KP, Sobral BW, Dickerman AW: **A robust species tree for the alphaproteobacteria.** *J Bacteriol* 2007, **189**:4578-4586.
37. Baldridge GD, Burkhardt NY, Felsheim RF, Kurtti TJ, Munderloh UG: **Transposon insertion reveals pRM, a plasmid of Rickettsia monacensis.** *Appl Environ Microbiol* 2007, **73**:4984-4995.
38. Baldridge GD, Burkhardt NY, Felsheim RF, Kurtti TJ, Munderloh UG: **Plasmids of the pRM/pRF Family Occur in Diverse Rickettsia species.** *Appl Environ Microbiol* 2008, **74**:645-652.
39. Matsunaga T, Okamura Y, Fukuda Y, Wahyudi AT, Murase Y, Takeyama H: **Complete genome sequence of the facultative anaerobic magnetotactic bacterium Magnetospirillum sp. strain AMB-1.** *DNA Res* 2005, **12**:157-166.
40. Burgdorfer W, Brinton LP: **Mechanisms of transovarial infection of spotted fever Rickettsiae in ticks.** *Ann N Y Acad Sci* 1975, **266**:61-72.
41. Cashel M, Gentry DR, Hernandez VJ: **The stringent response.** In *Escherichia coli and Salmonella: Cellular and Molecular Biology* Edited by: Neidhardt FC, Curtis III R, Ingraham JL, Lin ECC, Low KB, Magasanik B, et al. Washington, D.C.: ASM Press; 1996:1458-1496.
42. Rovery C, Renesto P, Crapoulet N, Matsumoto K, Parola P, Ogata H, *et al.*: **Transcriptional response of Rickettsia conorii exposed to temperature variation and stress starvation.** *Res Microbiol* 2005, **156**:211-218.
43. La MV, François P, Rovery C, Robineau S, Barbry P, Schrenzel J, *et al.*: **Development of a method for recovering rickettsial RNA from infected cells to analyze gene expression profiling of obligate intracellular bacteria.** *J Microbiol Meth* 2007, **71**:292-297.
44. Teysseire N, Raoult D: **Comparison of Western immunoblotting and microimmunofluoresence for diagnosis of Mediterranean spotted fever.** *J Clin Microbiol* 1992, **30**:455-460.
45. Li H, Walker DH: **RompA is a critical protein for the adhesion of *Rickettsia rickettsii* to host cells.** *Microbial Pathogenesis* 1998, **24**:289-298.
46. Martinez JJ, Seveau S, Veiga E, Matsuyama S, Cossart P: **Ku70, a component of DNA-dependent protein kinase, is a mammalian receptor for Rickettsia conorii.** *Cell* 2005, **123**:1013-1023.
47. Blanc G, Ngwamidiba M, Ogata H, Fournier PE, Claverie JM, Raoult D: **Molecular Evolution of Rickettsia Surface Antigens: Evidence of Positive Selection.** *Mol Biol Evol* 2005, **22**:2073-2083.
48. Walker TS, Winkler HH: **Penetration of cultured mouse fibroblasts (L cells) by *Rickettsia prowazeki*.** *Infect Immun* 1978, **22**:200-208.
49. Blanc G, Renesto P, Raoult D: **Phylogenic analysis of rickettsial patatin-like protein with conserved phospholipase A2 active sites.** *Ann N Y Acad Sci* 2005, **1063**:83-86.
50. Gouin E, Egile C, Dehoux P, Villiers V, Adams J, Gertler F, *et al.*: **The RickA protein of Rickettsia conorii activates the Arp2/3 complex.** *Nature* 2004, **427**:457-461.
51. Christie PJ, Vogel JP: **Bacterial type IV secretion: conjugation systems adapted to deliver effector molecules to host cells.** *Trends Microbiol* 2000, **8**:354-360.
52. Beati L, Kelly PJ, Matthewman LA, Mason PR, Raoult D: **Prevalence of rickettsia-like organisms and spotted fever group rickettsiae in ticks (Acari: Ixodidae) from Zimbabwe.** *J Med Entomol* 1995, **32**:787-792.
53. Fournier PE, Zhu Y, Ogata H, Raoult D: **Use of highly variable intergenic spacer sequences for multispacer typing of *Rickettsia conorii* strains.** *J Clin Microbiol* 2004, **42**:5757-5766.
54. Pallen MJ, Wren BW: **Bacterial pathogenomics.** *Nature* 2007, **449**:835-842.
55. Fraser-Liggett CM: **Insights on biology and evolution from microbial genome sequencing.** *Genome Res* 2005, **15**:1603-1610.
56. Raskin DM, Seshadri R, Pukatzki SU, Mekalanos JJ: **Bacterial genomics and pathogen evolution.** *Cell* 2006, **124**:703-714.
57. Fuxelius HH, Darby AC, Cho NH, Andersson SG: **Visualization of pseudogenes in intracellular bacteria reveals the different tracks to gene destruction.** *Genome Biol* 2008, **9**:R42.
58. Andersson SGE, Zomorodipour A, Andersson JO, Sicheritz-Pontén T, Alsmark UCM, Podowski RM, *et al.*: **The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria.** *Nature* 1998, **396**:133-140.
59. Walker DH, Yu XJ: **Progress in rickettsial genome analysis from pioneering of Rickettsia prowazekii to the recent Rickettsia typhi.** *Ann N Y Acad Sci* 2005, **1063**:13-25.
60. Zhang D, Xu Z, Sun W, Karaolis KR: **The *Vibrio* pathogenicity island-encoded Mop protein modulates the pathogenesis and reactogenicity of epidemic *Vibrio cholerae*.** *Infect immun* 2003, **71**:510-515.
61. Federle MJ, McIver KS, Scott JR: **A response regulator that represses transcription of several virulence operons in the group A streptococcus.** *J Bacteriol* 1999, **181**:3649-3657.
62. Glass JI, ssad-Garcia N, Alperovich N, Yooseph S, Lewis MR, Maruf M, *et al.*: **Essential genes of a minimal bacterium.** *Proc Natl Acad Sci USA* 2006, **103**:425-430.
63. Kaser M, Pluschke G: **Differential Gene Repertoire in Mycobacterium ulcerans Identifies Candidate Genes for Patho-Adaptation.** *PLoS Negl Trop Dis* 2008, **2**:e353.
64. Kelly PJ, Mason PR: **Transmission of a Spotted Fever Group Rickettsia by *Amblyomma hebraeum* (*Acari: Ixodidae*).** *J Med Entomol* 1991, **28**:598-600.
65. Burgdorfer W: **Investigation of transovarial transmission of *Rickettsia rickettsii* in the wood tick *Dermacentor andersoni*.** *Exp Parasitol* 1963, **14**:152.
66. Niebylski ML, Peacock MG, Schwan TG: **Lethal effect of Rickettsia rickettsii on its tick vector (Dermacentor andersoni).** *Appl Environ Microbiol* 1999, **65**:773-778.
67. Matsumoto K, Brouqui P, Raoult D, Parola P: **Experimental infection models of ticks of the Rhipicephalus sanguineus group with *Rickettsia conorii*.** *Vector Borne Zoonotic Dis* 2005, **5**:363-372.
68. Lepidi H, Fournier PE, Raoult D: **Histologic features and immunodetection of African tick-bite fever eschar.** *Emerg Infect Dis* 2006, **12**:1332-1337.
69. Eremeeva ME, Klemt RM, Santucci-Domotor LA, Silverman DJ, Dasch GA: **Genetic analysis of isolates of Rickettsia rickettsii that differ in virulence.** *Ann N Y Acad Sci* 2003, **990**:717-722.
70. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, *et al.*: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25**:3389-3402.
71. Bairoch A, Apweiler R, Wu CH, Barker WC, Boeckmann B, Ferro S, *et al.*: **The Universal Protein Resource (UniProt).** *Nucleic Acids Res* 2005, **33**:D154-D159.
72. Marchler-Bauer A, Anderson JB, Cherukuri PF, Weese-Scott C, Geer LY, Gwadz M, *et al.*: **CDD: a Conserved Domain Database for protein classification.** *Nucleic Acids Res* 2005, **33**:D192-D196.
73. Ponting CP, Schultz J, Milpetz F, Bork P: **SMART: identification and annotation of domains from signalling and extracellular protein sequences.** *Nucleic Acids Res* 1999, **27**:229-232.
74. Lowe TM, Eddy SR: **t-RNAscan-SE: a program for improved detection of transfer RNA gene in genomic sequence.** *Nucleic Acids Res* 1997, **25**:955-964.
75. Eddy SR: **Hidden Markov models.** *Curr Opin Struct Biol* 1996, **6**:361-365.
76. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG: **The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools.** *Nucleic Acids Res* 1997, **25**:4876-4882.
77. Notredame C, Higgins DG, Heringa J: **T-Coffee: A novel method for fast and accurate multiple sequence alignment.** *J Mol Biol* 2000, **302**:205-217.
78. Edgar RC: **MUSCLE: a multiple sequence alignment method with reduced time and space complexity.** *BMC Bioinformatics* 2004, **5**:113.
79. Katoh K, Misawa K, Kuma K, Miyata T: **MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform.** *Nucleic Acids Res* 2002, **30**:3059-3066.

80. Kumar S, Tamura K, Nei M: **MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment.** *Brief Bioinform* 2004, **5**:150-163.

Annexe 3

Genomic, proteomic, and transcriptomic analysis of virulent and avirulent *Rickettsia prowazekii* reveals its adaptive mutation capabilities

Yassina BECHAH[1], Khalid El KARKOURI[1], Oleg MEDIANNIKOV[1], Quentin LEROY[1], Nicolas PELLETIER[1], Catherine ROBERT[1], Claudine MEDIGUE[2], Jean-Louis MEGE[1], Didier RAOULT[1]

Genome Research

2010 May;20(5):655-63. Epub 2010 Apr 5

1 - Unité des Rickettsies, IFR 48 CNRS UMR 6020, Faculté de Médecine, Université de la Méditerranée, 27 Boulevard Jean Moulin, 13385 Marseille cedex 05, France

2 - Laboratoire d'Analyses Bioinformatiques pour la Génomique et le Métabolisme, CEA-Institut de Génomique, Génoscope & CNRS UMR 8030, 2 rue Gaston Crémieux, 91057 Evry cedex, France

Research

Genomic, proteomic, and transcriptomic analysis of virulent and avirulent *Rickettsia prowazekii* reveals its adaptive mutation capabilities

Yassina Bechah,[1] Khalid El Karkouri,[1] Oleg Mediannikov,[1] Quentin Leroy,[1] Nicolas Pelletier,[1] Catherine Robert,[1] Claudine Médigue,[2] Jean-Louis Mege,[1] and Didier Raoult[1,3]

[1]*Unit for Research on Emergent and Tropical Infectious Diseases (URMITE), Centre National de la Recherche Scientifique–Institut de Recherche pour le Développement, Faculty of Medicine, University of the Mediterranean, 13005 Marseille, France;* [2]*Genoscope, Centre National de Séquençage, Laboratoire de Génomique Comparative, 91057 Evry cedex, France*

Rickettsia prowazekii, the agent of epidemic typhus, is an obligate intracellular bacterium that is transmitted to human beings by the body louse. Several strains that differ considerably in virulence are recognized, but the genetic basis for these variations has remained unknown since the initial description of the avirulent vaccine strain nearly 70 yr ago. We use a recently developed murine model of epidemic typhus and transcriptomic, proteomic, and genetic techniques to identify the factors associated with virulence. We identified four phenotypes of *R. prowazekii* that differed in virulence, associated with the up-regulation of antiapoptotic genes or the interferon I pathway in the host cells. Transcriptional and proteomic analyses of *R. prowazekii* surface protein expression and protein methylation varied with virulence. By sequencing a virulent strain and using comparative genomics, we found hotspots of mutations in homopolymeric tracts of poly(A) and poly(T) in eight genes in an avirulent strain that split and inactivated these genes. These included *recO*, putative methyltransferase, and exported protein. Passage of the avirulent Madrid E strain in cells or in experimental animals was associated with a cascade of gene reactivations, beginning with *recO*, that restored the virulent phenotype. An area of genomic plasticity appears to determine virulence in *R. prowazekii* and represents an example of adaptive mutation for this pathogen.

[Supplemental material is available online at http://www.genome.org. The sequence data from this study have been submitted to GenBank (http://www.ncbi.nlm.nih.gov/Genbank) under accession no. CP001584. The microarray data from this study have been submitted to the NCBI Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo) under accession nos. GSE16123 and GSE15630.]

Epidemic typhus, caused by *Rickettsia prowazekii*, is a devastating disease with a mortality rate reaching 30% (Zinsser 1935; Bechah et al. 2008a). It has been reported to have killed millions of people during wars, including the Napoleonic (Raoult et al. 2006) and the First and the Second World Wars (Raoult and Roux 1999). The typical transmission of epidemic typhus involves human beings and lice (Raoult and Roux 1999). Epidemic typhus is characterized histologically by generalized vasculitis with increased vascular permeability, edema, mononuclear cell infiltrations, and activation of inflammatory mechanisms. *R. prowazekii* is known to target endothelial cells in vivo.

The prototypic virulent strain of *R. prowazekii* is Breinl (Ormsbee et al. 1978). Reference avirulent strain Madrid E was obtained after serial passages of virulent strain Madrid in eggs (Fox et al. 1954). It has been widely used as a vaccine (Fox et al. 1957). Its genome has been sequenced; it is small (1,111,523 bp) and contains large amounts (up to 24%) of noncoding DNA (Andersson et al. 1998). It was demonstrated that when Madrid E is cultivated in cell cultures (Ignatovich 1975) or when inoculated to animals (Balayeva and Nikolskaya 1972) instead of chick embryos, the virulence of bacterium is restored after several passages. This new strain that regained virulence (revertant strain) was named Evir (for virulent) (Balayeva and Nikolskaya 1972); however, no differences in protein pattern were detected when comparing parental Madrid E strain and Evir (Balayeva et al. 1992). In 1999, a new virulent strain of *R. prowazekii*, Rp22, was isolated from a patient (Birg et al. 1999).

Comparative genomic microarray study revealed highly conserved genome content between Breinl and Madrid E strains (only ~3% variation) (Ge et al. 2004).

Methylation has been proposed to play a major role in bacterial virulence. Indeed, the lysine methylation profile of surface proteins is different in Breinl and Evir compared with Madrid E (Rodionov et al. 1991; Ching et al. 1993; Turco and Winkler 1994). A methytransferase gene is inactivated in Madrid E (Zhang et al. 2006), and the *metK* gene, which codes for S-adenosylmethionine synthetase, is split in Madrid E but complete in Breinl (Andersson et al. 1998). Furthermore, restricted growth of Madrid E compared with Breinl and Evir is found in macrophages (Gambrill and Wisseman 1973; Turco and Winkler 1982).

Recently, we established an infection of BALB/c mice with Breinl that mimics human disease (Bechah et al. 2007). This murine model offers the opportunity to test the virulence of different *R. prowazekii* strains. This can also be assessed by increased migration of peripheral blood mononuclear cells (PBMCs) across endothelial cell monolayers (Bechah et al. 2008b). Here, we compare

[1]**Corresponding author.**
E-mail didier.raoult@medecine.univ-mrs.fr; fax 33-4-91-38-77-72.
Article published online before print. Article and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.103564.109.

the virulence phenotypes of various *R. prowazekii* strains with their genotypes and gene expression profiles by an integral approach to identify factors associated with virulence (Fig. 1).

Results

R. prowazekii phenotypes

We have succeeded in generation of a new strain of *R. prowazekii* from avirulent Madrid E-M (from Gamaleya Institute in Moscow, Russia). After 3 mo of culture on L929 cells, the growth rate of parental Madrid E strain has changed dramatically. The bacteria began to grow as fast as virulent strains, including Evir. Considering this change, we decided to call this hypothetically new strain Erus ("E" is for Madrid E, and "rus" is for Russia from which Madrid E came to our laboratory).

Having compared the properties of all strains, we found four different phenotypes for *R. prowazekii* strains (Supplemental Table 1). Two different phenotypes identified in cell culture Madrid E grew very slowly in cells compared with all other strains (Supplemental Table 1). A revertant (Erus) was obtained after 3 mo of culture of Madrid E in L929 cell line (Fig. 1), as previously reported (Ignatovich 1975). We found three levels of pathogenicity when inoculating mice using 10^5 *R. prowazekii* cells (Supplemental Table 1). Rp22, as was previously described for Breinl (Bechah et al. 2007), caused bacteremia for 10 d; its DNA was detected in the liver, lungs, and brain, and it caused histological lesions in these organs (Supplemental Fig. 1). Evir caused bacteremia, but no lesions were observed in the sampled organs. Erus did not cause bacteremia. After intradermic inoculation in guinea pigs with the same inocula, wild strains (Breinl, Rp22) and Evir caused erythematous skin lesions, but Erus did not. Histologically, lesions showed intense inflammatory infiltrate at the inoculum site consisting mainly of macrophages and lymphocytes (Supplemental Fig. 2). Finally, the infection of endothelial cells with Breinl, Rp22, and Evir, but not Erus, significantly ($P < 0.05$) increased leukocyte transmigration compared with uninfected endothelial cells (Supplemental Fig. 3). Our results show that Erus is avirulent and Evir is less pathogenic than wild strains, as proposed previously by Balayeva and Nikolskaya (1970).

Transcriptional profile of infected human endothelial cells

The total number of genes modulated in response to Rp22 ($n = 67$) or Erus ($n = 65$) infection is comparable (Fig. 2A). However, distinct clusters of genes were identified for each strain (Fig. 2B). Indeed, 16 genes were up-regulated by Erus (Supplemental Table 2A); seven genes were interferon 1 (IFN-I)-inducible. Eighteen genes were specifically up-regulated in response to Rp22, including eight associated with antiapoptotic responses (Supplemental Table 2B). An antiapoptotic profile was further confirmed as we found, by reverse transcription PCR, in which proapoptotic genes *BCL2*, caspase 8, and *NAIP* were markedly down-regulated in Rp22-infected cells (Supplemental Fig. 4). In contrast, Evir modulated the expression of only 30 genes. Among the modulated genes, there is induction of expression of *HAS1* (a gene involved in antiapoptotic response) as Rp22 and of *OASL* and *IFI44* (two genes of IFN-I-inducible gene pathway) as Erus.

R. prowazekii comparative genomics

We sequenced the genome of a virulent strain (Rp22) of *R. prowazekii*, and its main characteristics (GenBank accession no. CP001584) are shown in Table 1 and Figure 3. The genome of Rp22

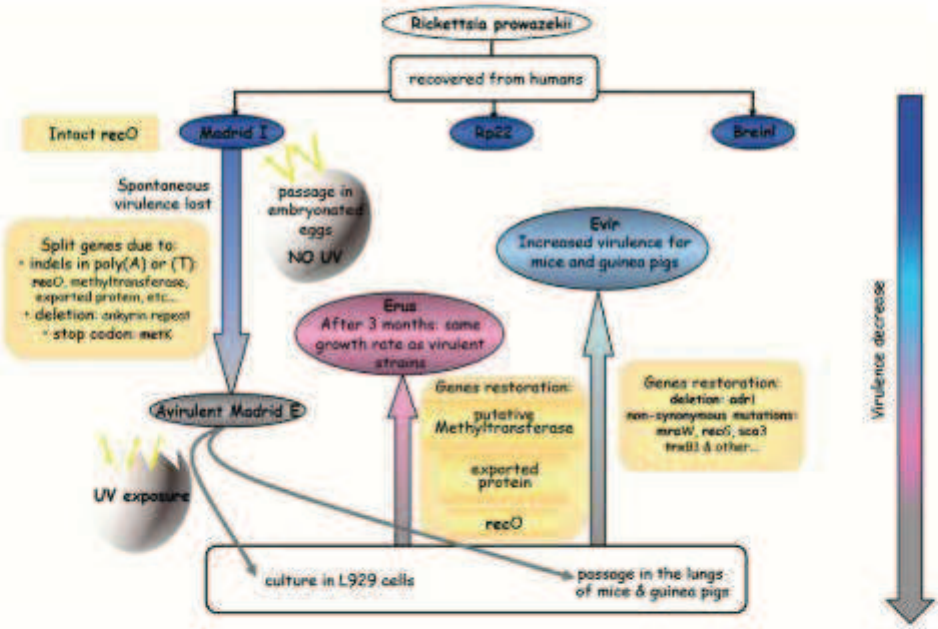


Figure 1. Scheme of *R. prowazekii* strains origin and evolution. The Breinl strain and the most recent isolate, Rp22, are considered highly virulent. The Madrid I strain was isolated in 1941 during the Madrid outbreak of epidemic typhus. After passages in embryonated eggs, Madrid I has lost its virulence and has been used under the name of Madrid E as a vaccine in humans since 1944. When it was inoculated to small rodents, Madrid E recovered some virulence (Evir). From Madrid E (300–600 passages in eggs), we have recently generated a new isolate by cultivating them in L 929 cells (Erus). (Breinl and Rp22) Virulent for humans and animals and replicates efficiently in L929 cells; (Evir) virulent for animals and replicates efficiently in L929 cells; (Erus) avirulent for animals, but replicates with L929; (Madrid E) avirulent for humans and animals and grows slowly in L929 cells.

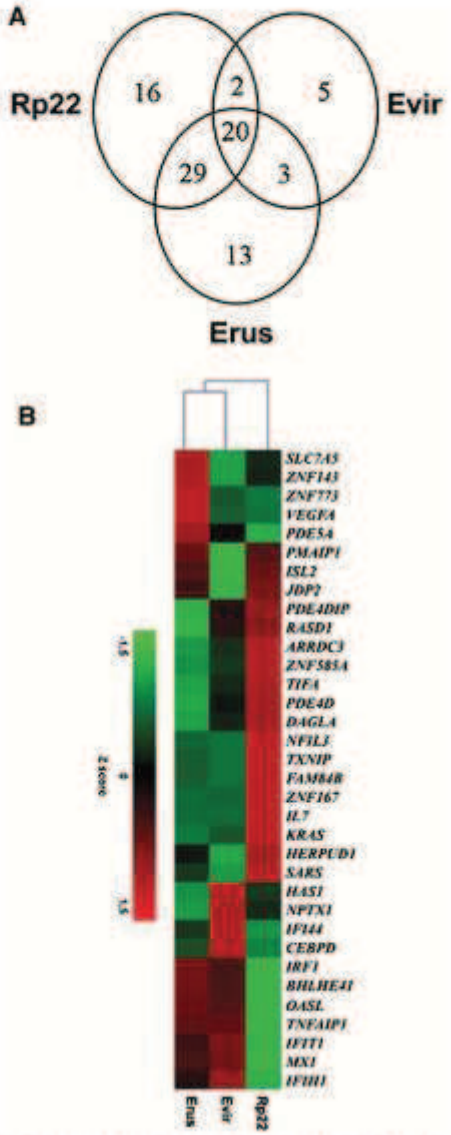


Figure 2. Transcriptional response of human endothelial cells infected by different strains of *R. prowazekii*. (*A*) Venn diagram illustrating the number of genes of endothelial cells differentially expressed in response to different *R. prowazekii* strains as compared with uninfected cells. (*B*) Transcriptional response analyzed by RNA microarrays. Modulated genes (fold change ≥ 2) were compared by unsupervised hierarchic clustering analysis.

consists of a single circular chromosome of 1,111,612 bp, with a G + C content of 29%. The RNAs from Rp22 and Madrid E exhibit identical sequences except for the 16S rDNA and a tRNA Val (Supplemental Fig. 5). We found only 81 genes differing between the two strains and classified them into four categories: 64 genes (including one split gene) with nonsynonymous mutations or gaps; 11 genes complete in Rp22 but split in Madrid E; four genes split in Rp22 but complete in Madrid E, and two genes split in both genomes at different places (Table 1). Indeed, significantly more genes are split in Madrid E (including *metK*, *recO*, and a putative methyltransferase) than in Rp22 (Table 2). Homopolymeric tracts of poly(A) and poly(T) were found significantly more often with split genes in both strains. Among the 15 genes split in either Rp22 or Madrid E, 10 showed insertion/deletion in homopolymeric poly(A) or poly(T) tracts of variable lengths (3–10 bp) (Supplemental Table 3; Fig. 4). The evolution of the Madrid E genome was found to be more rapid than that of Rp22. Using *R. typhi* as an outgroup, Madrid E showed a slightly higher K_a (nonsynonymous substitutions)/K_s (synonymous substitutions) ratio (0.378) and a longer phylogenetic branch length (0.00264) than Rp22 (0.336, 0.00122, respectively) (Supplemental Table 5; Supplemental Fig. 6).

The genes differing between these two strains were then amplified, sequenced, and compared with five additional *R. prowazekii* strains: Breinl, Madrid E USA (propagated in eggs in a laboratory in the Naval Medical Research Hospital), Madrid E-M (from the Gamaleya Institute), Evir, and Erus. This allowed the identification of four genotypes (Table 2) corresponding to the four different phenotypes. The Madrid E genotype possessed three genes differing from Erus. These genes may be responsible for phenotype differences between Erus and Madrid E in cell culture. A putative methyltransferase (RP027-28) was split in Madrid E and restored in Erus. This gene has been proposed as a candidate gene for the difference of virulence between Madrid E and Evir (Zhang et al. 2006) . However, Erus is not pathogenic; therefore, we conclude that the presence of this gene did not fully restore the virulence. A gene that repairs DNA, *recO* (RP548m-48m), was also restored in Erus. The third restored gene (RP061) had no identified function. Interestingly, early sampling during Erus transition from Madrid E showed that *recO* and Rp061 were the first restored genes (data not shown).

In a second step, we compared Erus with less pathogenic Evir and with virulent strains. No split gene differences were found; however, we found a deletion in the adhesin gene *adr1* (Table 2). We also found nonsynonymous mutations in 14 genes (Supplemental Table 6). They include surface protein (Sca3), thioredoxin (TrxB1), and genes coding for a manosyltransferase (RP340) and a putative acyltransferase (RP804). Finally, in a third step of comparing less pathogenic Evir and wild strains, we found six split genes in Evir (Table 2), including *metK* (S-adenosylmethionine synthetase) and an ankyrin repeat-containing protein. We found an insertion in *asmA* (coding for outer membrane protein assembly) in avirulent and less pathogenic strains. We also identified nonsynonymous mutations in several genes, including those coding for surface proteins of the Sca family (Sca1, Sca4, Sca5, and Sca6).

R. prowazekii gene expression

Transcriptome analysis revealed that 77 genes were differentially expressed in Rp22, as compared with Erus, when cultured in L929 cells (Fig. 5). Functional category classification identified genes

Table 1. Comparison analysis of wild strain (Rp22) and vaccine-attenuated strain (Madrid E) genomes

Species	Rp22	Madrid E
Genome size	1,111,612	1,111,523
G + C content (%)	29	29
Genes	966	966
tRNAs	33	33
rRNAs	3	3
Other RNAs	3	3
Complete and conserved genes	878	878
Nonsynonymous mutation or gapped genes	64	64
Split in one strain only	4	11
Split (same) in both strains	7	7
Split (different) in both strains	2	2

PCRs were applied for genes that were found to be different between the two sequence genomes.

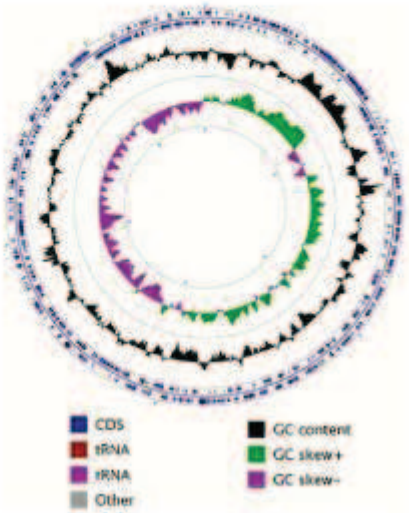


Figure 3. Circular representation of *R. prowazekii* Rp22 genome. The circles show the G-C content and skews, CDSs, rRNA, tRNA, and miscellaneous RNA (other).

mainly implicated in translation, replication, cell wall/membrane biogenesis, and post-transcriptional modifications. Thirty-three genes were classified as function unknown, general function prediction only, or not found in the COG (Clusters of Orthologous Groups) database (Supplemental Table 7). Three genes (*rOmpB*, mannosytransferase, and *asmA*) with expression differences between Rp22 and Erus had different genetic sequences (non-synonymous mutations for the two first genes and 18 coding base pair insertions for *asmA*) (Supplemental Table 6). These three genes are implicated in cell wall membrane biogenesis and are down-regulated in Rp22. In addition, Rp22 up-regulates the expression of an ankyrin repeat-containing protein, a heat shock protein (HtpG), and a peptidoglycan-associated lipoprotein precursor. The increased production of the latter was further confirmed.

Proteomic analysis showed that six proteins were down-regulated in Erus compared with Rp22 (Supplemental Table 8). The six proteins included two putative methyltransferases, translation-associated GTPase, peptidoglycan-associated lipoprotein precursor, heat shock protein (Hsp22), and preprotein translocase. Non-synonymous mutations were identified in genes coding the first three proteins. In contrast to Erus, Rp22 down-regulated the protein thioredoxin reductase (TrxB1) (Supplemental Table 8). However, this experimental result should be considered carefully, because the *trx*B1 gene has a nonsynonymous mutation (E/K) that can result in the shift of pI in a two-dimensional gel. In addition, MetK protein was produced only in Rp22 as predicted by genomic study (Table 2). Lysine methylation in both strains was also studied (Fig. 6). Surprisingly, the methylation of lysine in Erus was closer to that of Rp22 (Fig. 6A–D) than to Madrid E (Rodionov et al. 1991; Ching et al. 1993; Turco and Winkler 1994). Among all proteins only putative methyltransferase (RP789) was found in four lysine methylated isoforms in Rp22 (Fig. 6C,E) and in two isoforms in Erus (Fig. 6D,F).

Altogether, transcriptomic and proteomic analysis revealed that major differences concerned surface-exposed proteins, post-translational modification (methylation), and modifications associated with modulation of stress response. *R. prowazekii* modulated its surface proteins to escape immune recognition.

Table 2. Summary of the differences among four *R. prowazekii* genotypes

RP22 ID	ME MaGe	COG functional categories	Annotation	Brein, Rp22[a]	Evir	Erus	ME	Rp22/ME
rpr22_CDS759	RP777-777	Coenzyme metabolism	*metK*; S-adenosylmethionine synthetase	TAT	TAG	TAG	TAG	Complete/split
rpr22_CDS740	RP757-RICPRO828	Function unknown	Unknown	—	A	A	A	Complete/split
rpr22_CDS769	RP787-88	Not in COGs	Similarity to putative invasin, adhesin and agglutinin 1	I	D	D	D	Complete/split
RPR22_0486	n.i.[b]	Not in COGs	Hypothetical protein	A	—	—	—	Complete/split
rpr22_CDS010	RP012	Not in COGs	Similarity to polysaccharide deacetylase precursor	I	D	D	D	Complete/split
rpr22_CDS699	RP716	Not in COGs	Ankyrin repeat	I	D	D	D	Complete/split
rpr22_CDSx325	RICPRO325	Not in COGs	ABC transporter ATP-binding protein	I	D	D	D	Split/complete
rpr22_CDSx885	RICPRO885	Not in COGs	Hypothetical protein	—	A	A	A	Split/complete
rpr22_CDS340	RP347	Cell envelope biogenesis, outer membrane	*asmA*; Outer membrane assembly protein	D	I	I	I	−/ANQKPK
rpr22_CDS551	RP573	Function unknown	Exported protein of unknown function	D	I	I	I	−/NQ
rpr22_CDS808	RP827	Cell envelope biogenesis, outer membrane	*adr1*; *Rickettsia* adhesin	I	I	D	D	TKTI/−
rpr22_CDS024	RP027-28	General function prediction only	Putative methyltransferase	—	—	—	A	Complete/split
rpr22_CDS529	RP548m-48m	DNA replication, recombination, and repair	*recO*; DNA repair protein	—	—	—	T	Complete/split
rpr22_CDS059	RP061	Not in COGs	Exported protein of unknown function	—	—	—	T	Complete/split

(ME) Madrid E; (I) insertion of bases; (D) deletion of bases.
[a]Rp22 and Breinl yield the same results.
[b]CDSs not identified in MaGe and/or RickBase, but found to be present in *R. prowazekii* genomes.

Figure 4. Sequence alignments of three *R. prowazekii* genes compared between Rp22 (rpr22) and Madrid E (rpr) strains. Split genes in Madrid E were due to stop codon (*A*), poly(A) (*B*), or poly(T) (*C*).

Discussion

There are no genetic manipulation techniques currently available for *Rickettsia*, which makes the study of its virulence difficult. *R. prowazekii* strains exhibit various degree of virulence (Turco and Winkler 1982), and a integral approach may help to better understand mutations involved in the different steps of virulence (Renesto et al. 2005). On the basis of our experimental models, we identified four different phenotypes (Fig. 1; Supplemental Table 1). This was critical, as Evir has been widely considered to be a fully restored virulent mutant (Balayeva and Nikolskaya 1973). Moreover, we confirmed that Madrid E grown in L929 cells easily generates a revertant.

To survive in a human host, *R. prowazekii* needs to avoid destruction and premature death. We found that the virulent strain inhibits IFN type I-inducible genes and promotes antiapoptotic genes to prevent death. Manipulation of IFN type I has been reported for viruses (Kato et al. 2006) and for bacteria surviving in the cytosol of infected cells such as *Listeria monocytogenes* (Leber et al. 2008). The control of apoptosis by bacteria of the order *Rickettsiales* has already been documented (Ge and Rikihisa 2006). Host manipulation by *R. prowazekii* may be mediated by ankyrin repeat-containing proteins that were considered to be key factors of host manipulation in other intracellular bacteria (Pan et al. 2008). Post-translational modifications may also play a role in the host–*R. prowazekii* relationship as was shown previously in other host–parasite systems (Polevoda and Sherman 2007). The relation between post-translational modifications, especially methylation, and virulence of *R. prowazekii* strains has been previously suggested (Rodionov et al. 1991; Ching et al. 1993; Turco and Winkler 1994; Chao et al. 2004). Here, we report only subtle differences in methylation of proteins, mainly concerning methyltransferase (RP789) itself, when comparing Rp22 with Erus. However, proteomic analysis showed that both methyltransferases (RP789 and RP527) are overproduced by Rp22, distinguishing it from Erus. Altogether, our data confirm that methylation plays a role in the virulence of *R. prowazekii* but that incompletely explains the loss or gain of virulence.

Surface-exposed proteins also seem to play a role in the variation of *R. prowazekii* virulence. The Sca family of proteins, including rOmpA (Sca0), rOmpB (Sca5), and the recently identified adhesins (Adr1 and Adr2), are major surface-exposed proteins critical in immunity (Blanc et al. 2005) and *Rickettsia*-eukaryotic cell interaction (Pizarro-Cerda and Cossart 2006; Renesto et al. 2006). We found multiple nonsynonymous mutations in *sca* family genes that possibly affect their functions among *R. prowazekii* strains. In addition, we found that the *adr1* gene was conserved among virulent and less pathogenic strains but altered (with a deletion) in Madrid E.

We identified three stages in genome degradation and restoration. Surprisingly, several genes split in the vaccine strain were able to revert to the wild type, confirming that this process is reversible. Gene repair in this process may be dependent on *recO* reparation, as it is the first identified step associated with virulence reversion. This gene codes for a protein involved in a protein complex (RecFOR) that plays a major role in DNA reparation (Chow and Courcelle 2004). We believe that its inactivation may act as a trigger in the loss of virulence of *R. prowazekii* when grown in eggs and that its inactivation favors the rapid alteration of other genes. This type of gene inactivation has been previously described in mutator clones (Caporale 2003).

The RecFOR epistatic group of proteins is implicated in the reparation of ultraviolet (UV)-damaged DNA (Chow and Courcelle 2004). Knock-out of a single coding gene may significantly complicate or even block DNA replication. We hypothesize that when rickettsiae systematically grew in chick embryos, the nearly complete absence of UV radiation and following relatively low amount of UV-damaged DNA make *recO* underused. Accidental knockout of a *recO* gene that usually may be lethal is tolerated in this situation. The beginning of cultivation of Madrid E in cell lines or animals, where UV exposure is significantly higher, will select for the rare occasional revertants (with intact *recO*). Our hypothesis is that the UV-based negative selection may play a role in case of restoration of multiple damaged genes in *R. prowazekii* via reactivation of the RecFOR system.

We believe that this may be relevant to the selection of *R. prowazekii* vaccine strain (Madrid E) in embryonated eggs, as this process has been considered an irreversible step to genome degradation (Dale et al. 2003; Lescot et al. 2008). In our work we found that, surprisingly, the restoration of *recO* gene was rapidly followed by the restoration of other genes. Restored genes, including those coding for post-translational modifications, may play a role in pathogenicity. The reparation of degraded genes after *recO* restoration may explain how the virulence may be restored in several steps.

We believe that we identified hotspot zones of plasticity in the genome of *R. prowazekii*. Poly(A) are common among AT-rich endosymbionts, and polymerase infidelity has been proved to impair and rescue gene functions (Tamas et al. 2008). We observed that revertants can be easily obtained when genes are split because of indels in poly(A) tracts. In *R. prowazekii*, three genes that are critical for pathogenicity in humans and mice were split because of this phenomenon. The reversion phenomenon is easily induced, and this explains why mutants and revertants have been produced in vitro. On the basis of our data, we think that these mutations

Figure 5. Transcriptional profile of Rp22 vs. Erus cultured on L929 cells and analyzed by RNA microarray. Regulated genes (fold change ≥ 1.5) in Rp22 as compared with Erus were represented using cluster analysis software.

resulting in a premature stop codon are unstable, and we prefer to use the name "split gene" instead of the previously used terms "pseudogene" or "pseudo-pseudogenes" (Baranov et al. 2005). *R. prowazekii* therefore is exemplifying the adaptive mutation concept (Rosenberg 2001), i.e., mutations formed in response to an environment. The reversal mutation of the *recO* gene occurring in a hotspot is rapidly followed by a cascade of gene restoration.

In conclusion, we found that several genes are involved in virulence differences and are mainly associated with surface proteins and post-translational modification. A key factor in the generation of the avirulent mutant is the knock out of the *recO* gene, a homopolymeric tract that is critical in accelerating genome degradation generating a mutator clone because surprisingly, its restoration permitted the restoration of putative pseudogenes.

Methods

R. prowazekii preparation

The strains used in this study are described in Figure 1, including Breinl strain; Rp22 strain, a clinical isolate that had 10 passages on L929 cells; and Evir strain, a putative virulent revertant of Madrid E selected in mice. Initially grown in embryonated eggs (300–600 passages), the vaccine Madrid E strain was cultured for 90 d in L929 cells. Bacteria grew very slowly in the cell culture, so only two passages have been performed. When, after three months, bacterial replication reached the growth rate of other strains, we called this strain "Erus."

R. prowazekii strains were propagated in L929 cells as previously described (Bechah et al. 2007). For in vivo and genomic studies, bacteria were harvested, pelleted, and purified by renographin density gradient centrifugation. For transcriptomic study, infected cells were lysed with 1% trypsin, and cell debris were eliminated by successive centrifugations at 110*g*. Bacteria were stored at −80°C before RNA extraction. For proteomic study, infected cells were lysed by sonication, and the bacterial pellet was purified in a sucrose gradient and a discontinuous renographin gradient. Bacteria were sonicated, and protein extracts were precipitated using the PlusOne 2-D Clean-Up Kit as described previously (Kowalczewska et al. 2006). They were stored at −80°C until cyanine labeling and/or isoelectric focusing (IEF).

Animal infection

BALB/c mice (7-wk-old females) were inoculated intravenously with different strains of *R. prowazekii*. For each strain, eight mice

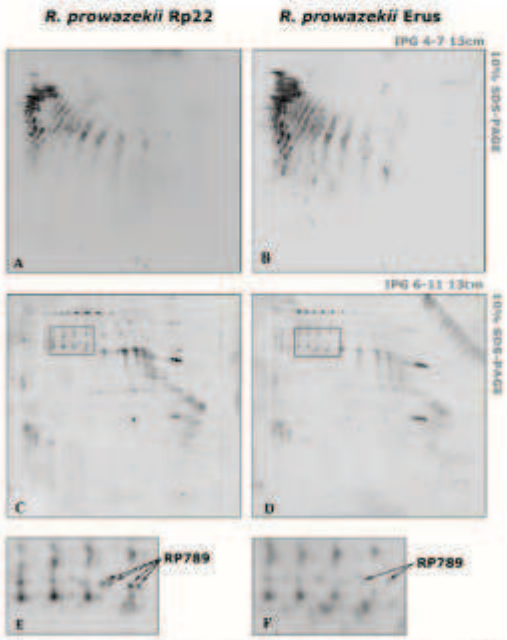


Figure 6. Lysine methylation in Rp22 (*A*,*C*,*E*) and Erus (*B*,*D*,*F*) strains. Two-dimensional Western blots were performed, and lysine methylation was visualized in pH range 4.0–7.0 (*A*,*B*) and pH range 6.0–11.0 (*C*,*D*). The zone that differed among both bacterial strains is boxed in *C* and *D*. Arrows in *E* and *F* indicate methylated spots of Rp22 and Erus strains that were identified as putative methyltransferase (RP789).

were inoculated intravenously with 5×10^5 bacteria and eight mice with 10^5 bacteria, as recently described (Bechah et al. 2007). Eight mice were also injected with 1.5×10^6 Erus bacteria/mouse. The clinical status of mice was recorded twice daily. Blood was collected and stored at −20°C for PCR. Mice were sacrificed at days 4, 6, and 10 post-infection (p.i.). Livers, lungs, and brains were aseptically excised. Half of the samples were stored at −80°C (for PCR) while the rest were fixed in 4% formalin and then embedded in paraffin for histological studies.

Six-week-old female Hartley guinea pigs were inoculated intradermally with 10^6 bacteria of each strain after shaving in separate back areas. The apparition of inflammatory lesions, which were defined as erythematous lesions without ulceration or necrosis of a size >1 mm and <2 mm, was recorded daily. A cutaneous biopsy was taken 7 d p.i. for histological examination. The 3-μm sections of paraffin-embedded specimens from guinea pigs or mice were stained with hematoxylin and eosin stain.

Bacteremia and tissue infection were determined as previously described (Bechah et al. 2007).

PBMC migration across endothelial cells

Confluent cells of the murine lung microvascular EC line L2 were seeded in gelatin-coated inserts (8-μm pore size, Costar) in 24-well plates (5×10^4 cells per well) and cultured until tight confluence. They were then infected with *R. prowazekii* organisms (bacterium-to-cell ratio of 50:1) for 6 h, as recently reported (Bechah et al. 2008b). Murine PBMCs were added to infected endothelial cells (4×10^5/well) for 24 h, and the percentage of PBMCs that had migrated across endothelial cell monolayers was determined by optical enumeration.

Microarrays of endothelial cells

Human endothelial cells (HMEC-1 cell line) were infected with *R. prowazekii* (50:1 bacterium-to-cell ratio) for 6 h at 37°C. Total RNA was extracted using an RNeasy Mini kit (Qiagen). RNA (300 ng) was reverse transcribed into cDNA with the M-MLV reverse transcriptase (RT). Labeled cRNA was synthesized from cDNA using T7 RNA polymerase and cyanine 3-labeled CTP (Cy-3) fluorescent dyes with the One-Color Low RNA Input Linear Amplification KitPLUS (Agilent). Fluorescent cRNAs were fragmented and hybridized onto 4×44k human wholegenome microarrays (Agilent) for 17 h at 65°C (Agilent). Slides were scanned at a 5-μm resolution with a G2505B DNA microarray scanner (Agilent). Agilent Feature Extractor Software A.9.1.3 was used for image analysis. The microarray grid was automatically placed, outlier spots were flagged, and feature intensities and backgrounds were accurately determined. Further data processing was performed using Resolver software 7.1 (Rosetta Inpharmatics), and its error model-based transformation pipeline was used to map replicate reporters to genes and to normalize inter-array data sets. The fold change (FC) was calculated using pairs Erus vs. Control, Evir vs. Control, and Rp22 vs. Control. Only genes that had an absolute FC over 2.0 ($P < 0.01$) for at least one of them were considered to be regulated and kept for further statistical analysis. Gene families were determined using numerous databases, including SOURCE (http://smd.stanford.edu/cgi-bin/source/sourceSearch) and Babelomics Fatigo+ (http://babelomics2.bioinfo.cipf.es/fatigoplus/cgi-bin/fatigoplus.cgi). The data have been submitted to the NCBI Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/) under accession numbers GPL4133 and GSE16123 for the platform of the microarray and the experimental data set, respectively.

Genome sequencing, annotation, and comparison of *R. prowazekii* strains

The genome of the Rp22 strain was sequenced using the whole-genome shotgun approach with a final coverage of 12× (Sanger method) as previously reported (Raoult et al. 2003). Genomic sequences were then assembled into contigs using *phred/phrap/consed* (Gordon et al. 1998) (http://www.phrap.org/phredphrapconsed.html), and all gaps were closed using PCR amplification and sequencing with specifically designed primers. Potential coding sequences (CDSs) were predicted using AMIGene (http://www.genoscope.cns.fr/agc/tools/amigene/Form/form.php). Split genes (either frameshifts with an in-frame stop codon or a change of their putative initiation/termination codon) or nonpredicted genes were detected and corrected manually where appropriate using Artemis (http://www.sanger.ac.uk/Software/Artemis/), BLASTN, and/or NUCmer from the MUMmer package. Assignment of protein functions was performed by searching the RickBase and MaGe databases (Vallenet et al. 2006; Blanc et al. 2007) using BLASTP. The COG functional categories of proteins were assigned using COGnitor (Tatusov et al. 1997). Ribosomal RNAs, tRNAs, and other RNAs were identified using BLASTN or tRNAscan-SE.

The Rp22 genome was then compared with that of *R. prowazekii* strain Madrid E (GenBank accession number NC_000963). Reciprocal-best BLAST matching was used to compare CDSs, proteins, and RNAs. Genes that were conserved, with or without synonymous mutations, were distinguished from those exhibiting nonsynonymous mutations, insertion/deletion(s), and/or a split state. Subsequently, discriminatory genes were compared among the *R. prowazekii* strains Breinl, Evir, Erus, Madrid E USA (propagated in eggs in a laboratory in the Naval Medical Research Hospital), and Madrid E-M (from Gamaleya Institute) and aligned using CLUSTALW. For each gene, differences were checked using amplification and sequencing with primers targeting flanking sequences.

The K_a/K_s ratio was used to estimate genomic divergence between *R. prowazekii* strains Rp22 and ME. The pairwise K_a/K_s ratios between *R. prowazekii* strains and *R. typhi* were obtained from the concatenated nucleotide alignment of 44 orthologous protein-coding genes using the K-Estimator software (Comeron 1999). The amino acid sequences of these 44 orthologs were concatenated for each genome, and multiple alignment was performed using Mafft software (Katoh et al. 2002). This was used to infer a Neighbor Joining tree using the MEGA 3.1 software (Kumar et al. 2004).

Study of bacterial transcriptome

Bacterial RNA was extracted using lysozyme and the RNeasy Mini kit (Qiagen) as recommended by the manufacturer, and DNA was digested by DNase. cDNA was synthesized from 500 ng of RNA with a random primer and the M-MLV RT. cDNA from Erus was used as reference DNA, whereas the cDNA from Rp22 strain was referred to as test DNA. Reference cDNA was amplified and labeled with cyanine 3-labeled CTP (Cy3) using the Bioprime DNA Labeling System (Invitrogen), whereas test cDNAs were labeled with cyanine 5-CTP (Cy5). Fluorescent cDNAs were fragmented and hybridized, as described above, onto *Rickettsia* Genome microarrays (Agilent Technologies) comprising probes specific for all genes and spacers from *R. prowazekii* strain Madrid E. Slides were scanned with XDR range at a 5-μm resolution. Feature Extractor 7.1 (Agilent Technologies) was used for image analysis. Data filtering and normalization were then performed using the Midas module of TM4 (La et al. 2007). The background-subtracted signals were normalized by the local subtraction method, and intensity signals were normalized by the global lowess method (La et al.

2007). Normalized signals were used for analysis with the TMev module of TM4. A *t*-test based on *P*-value permutation with a Bonferroni correction was used. Only genes with a fold change >1.5 ($P < 0.05$) were considered to have significant differential expression. The data have been submitted to the NCBI GEO (http://www.ncbi.nlm.nih.gov/geo/) under accession numbers GPL8427 and GSE15630 for the platform of the microarray and the experimental data set, respectively.

Real time RT-PCR

qRT-PCR was performed as previously described (Bechah et al. 2008b). The primer sequences are listed in Supplemental Table 4. Reverse transcriptase was omitted in the negative control. The FC in target gene cDNA relative to the housekeeping gene was determined by the $2^{-\Delta\Delta C_T}$ method (Bechah et al. 2008b). Only gene expression with a FC of >1.5 was considered to be modulated.

Differential gel electrophoresis

Differential gel electrophoresis (DIGE) was performed as follows: Each protein sample (50 μg) was labeled with 400 pmol of Cy3 or Cy5, and Cy2 was used as internal calibrator as previously described (Alban et al. 2003). IEF was performed according to the manufacturer's protocol (Ettan IPGphor II, GE Healthcare) using two pH ranges (4–7 and 6–11) of Immobiline DryStrips. For the second dimension, proteins were resolved by 10% SDS polyacrylamide gel electrophoresis (SDS-PAGE) (Ettan IPGphor II, GE Healthcare), and the gels were digitized using a Typhoon Trio scanner (Ettan IPGphor II, GE Healthcare) at wavelengths of 532 nm for Cy3, 633 nm for Cy5, and 488 nm for Cy2, and scanned at a 100-μm resolution. Protein quantification and statistical analysis were carried out using the Decyder software program (GE Healthcare). Spots with twofold or greater changes were considered significantly different and excised for identification by mass analyses (MALDI-TOF/TOF Bruker Ultraflex II spectrometer, Bruker Daltonics).

Methylation studies

For methylation studies, resolved proteins were transferred onto nitrocellulose membranes (Bio-Rad) using a semidry transfer unit (Hoefer Scientific). Membranes were then blocked in TBS (20 mM Tris-HCl at pH 7.5, 150 mM NaCl) supplemented with 0.1% Tween-20 and 5% bovine serum albumin for 1.5 h before incubation with rabbit antibodies directed against methyl-lysine (Biomol GmbH) diluted at 1:400. After 1 h, membranes were washed and probed with 1:1,000 horseradish peroxidase-conjugated goat anti-rabbit IgG (GE Healthcare). Spots were visualized using an ECL kit (GE Healthcare). Reactive spots were identified as above.

Acknowledgments

We thank R. Allen for providing us with DNA of the Madrid E strain, I. Tarasevitch for giving us the Evir strain, and S. Audic for his valuable comments. We also thank S. Azza, Y. Lepriol, T. N'Guyen, and H. Parrinello for technical assistance in proteome, transcriptome, and sequencing experiments. We thank Isabelle Combe for helping in formatting the manuscript.

References

Alban A, David SO, Bjorkesten L, Andersson C, Sloge E, Lewis S, Currie I. 2003. A novel experimental design for comparative two-dimensional gel analysis: Two-dimensional difference gel electrophoresis incorporating a pooled internal standard. *Proteomics* **3:** 36–44.

Andersson SGE, Zomorodipour A, Andersson JO, Sicheritz-Pontén T, Alsmark UCM, Podowski RM, Näsland AK, Eriksson AS, Winkler HH, Kurland CG. 1998. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* **396:** 133–140.

Balayeva NM, Nikolskaya VN. 1970. Stability of increased virulence of vaccine strain E of *Rickettsia prowazekii*. *J Hyg Epidemiol Microbiol Immunol* **8:** 36–38.

Balayeva NM, Nikolskaya VN. 1972. Enhanced virulence of the vaccine strain E of *Rickettsia prowazeki* on passaging in white mice and guinea pigs. *Acta Virol* **16:** 80–82.

Balayeva NM, Nikolskaya VN. 1973. Analysis of lung culture of *Rickettsia prowazeki* E strain with regard to its capacity of increasing virulence in passages on the lungs of white mice. *J Hyg Epidemiol Microbiol Immunol* **17:** 294–303.

Balayeva NM, Eremeeva ME, Ignatovich VF, Dmitriev BA, Lapina EB, Belousova IS. 1992. Protein antigens of genetically related *Rickettsia prowazekii* strains with different virulence. *Acta Virol* **36:** 52–56.

Baranov PV, Hammer AW, Zhou J, Gesteland RF, Atkins JF. 2005. Transcriptional slippage in bacteria: Distribution in sequenced genomes and utilization in IS element gene expression. *Genome Biol* **6:** R25. doi: 10.1186/gb-2005-6-3-r25.

Bechah Y, Capo C, Grau GE, Raoult D, Mege JL. 2007. A murine model of infection with *Rickettsia prowazekii*: Implications for pathogenesis of epidemic typhus. *Microbes Infect* **9:** 898–906.

Bechah Y, Capo C, Mege JL, Raoult D. 2008a. Epidemic typhus. *Lancet Infect Dis* **8:** 417–426.

Bechah Y, Capo C, Raoult D, Mege JL. 2008b. Infection of endothelial cells with virulent *Rickettsia prowazekii* increases the transmigration of leukocytes. *J Infect Dis* **197:** 142–147.

Birg ML, La Scola B, Roux V, Brouqui P, Raoult D. 1999. Isolation of *Rickettsia prowazekii* from blood by shell vial cell culture. *J Clin Microbiol* **37:** 3722–3724.

Blanc G, Ngwamidiba M, Ogata H, Fournier PE, Claverie JM, Raoult D. 2005. Molecular evolution of *Rickettsia* surface antigens: Evidence of positive selection. *Mol Biol Evol* **22:** 2073–2083.

Blanc G, Ogata H, Robert C, Audic S, Suhre K, Vestris G, Claverie JM, Raoult D. 2007. Reductive genome evolution from the mother of Rickettsia. *PLoS Genet* **3:** e14. doi: 10.1371/journal.pgen.0030014.

Caporale LH. 2003. Natural selection and the emergence of a mutation phenotype: An update of the evolutionary synthesis considering mechanisms that affect genome variation. *Annu Rev Microbiol* **57:** 467–485.

Chao CC, Wu SL, Ching WM. 2004. Using LC-MS with de novo software to fully characterize the multiple methylations of lysine residues in a recombinant fragment of an outer membrane protein from a virulent strain of *Rickettsia prowazekii*. *Biochim Biophys Acta* **1702:** 145–152.

Ching WM, Wang H, Davis J, Dasch GA. 1993. Amino acid analysis and multiple methylation of lysine residues in the surface protein antigen of *Rickettsia prowazekii*. In *Techniques in Protein Chemistry* (ed. R. Angeletti), pp. 307–314. Academic Press, San Diego, CA.

Chow KH, Courcelle J. 2004. RecO acts with RecF and RecR to protect and maintain replication forks blocked by UV-induced DNA damage in *Escherichia coli*. *J Biol Chem* **279:** 3492–3496.

Comeron JM. 1999. K-Estimator: Calculation of the number of nucleotide substitutions per site and the confidence intervals. *Bioinformatics* **15:** 763–764.

Dale C, Wang B, Moran N, Ochman H. 2003. Loss of DNA recombinational repair enzymes in the initial stages of genome degeneration. *Mol Biol Evol* **20:** 1188–1194.

Fox JP, Everritt MG, Robinson TA, Conwell DP. 1954. Immunization of man against epidemic typhus by infection with avirulent *Rickettsia prowazeki* (strain E); observations as to post-vaccination reactions, the relation of serologic response to size and route of infecting dose, and the resistance to challenge with virulent typhus strains. *Am J Hyg* **59:** 74–88.

Fox JP, Jordan ME, Gelfand HM. 1957. Immunization of man against epidemic typhus by infection with avirulent *Rickettsia prowazeki* strain E. IV. Persistence of immunity and a note as to differing complement-fixation antigen requirements in post-infection and post-vaccination sera. *J Immunol* **79:** 348–354.

Gambrill MR, Wisseman CL Jr. 1973. Mechanisms of immunity in typhus infections. I. Multiplication of typhus rickettsiae in human macrophage cell cultures in the nonimmune system: Influence of virulence of rickettsial strains and of chloramphenicol. *Infect Immun* **8:** 519–527.

Ge Y, Rikihisa Y. 2006. *Anaplasma phagocytophilum* delays spontaneous human neutrophil apoptosis by modulation of multiple apoptotic pathways. *Cell Microbiol* **8:** 1406–1416.

Ge H, Chuang YY, Zhao S, Tong M, Tsai MH, Temenak JJ, Richards AL, Ching WM. 2004. Comparative genomics of *Rickettsia prowazekii* Madrid E and Breinl strains. *J Bacteriol* **186:** 556–565.

Gordon D, Abajian C, Green P. 1998. *Consed*: A graphical tool for sequence finishing. *Genome Res* **8:** 195–202.

Ignatovich VF. 1975. Enhancement of the antigenic activity and virulence of the vaccine strain E of *Rickettsia prowazeki* by passages in cell culture. *Acta Virol* **19:** 481–485.
Kato H, Takeuchi O, Sato S, Yoneyama M, Yamamoto M, Matsui K, Uematsu S, Jung A, Kawai T, Ishii KJ, et al. 2006. Differential roles of MDA5 and RIG-I helicases in the recognition of RNA viruses. *Nature* **441:** 101–105.
Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res* **30:** 3059–3066.
Kowalczewska M, Fenollar F, Lafitte D, Raoult D. 2006. Identification of candidate antigen in Whipple's disease using a serological proteomic approach. *Proteomics* **6:** 3294–3305.
Kumar S, Tamura K, Nei M. 2004. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform* **5:** 150–163.
La MV, Crapoulet N, Barbry P, Raoult D, Renesto P. 2007. Comparative genomic analysis of *Tropheryma whipplei* strains reveals that diversity among clinical isolates is mainly related to the WiSP proteins. *BMC Genomics* **8:** 349. doi: 10.1186/1471-2164-8-349.
Leber JH, Crimmins GT, Raghavan S, Meyer-Morse NP, Cox JS, Portnoy DA. 2008. Distinct TLR- and NLR-mediated transcriptional responses to an intracellular pathogen. *PLoS Pathog* **4:** e6. doi: 10.1371/journal.ppat.0040006.
Lescot M, Audic S, Robert C, Nguyen TT, Blanc G, Cutler SJ, Wincker P, Couloux A, Claverie JM, Raoult D, et al. 2008. The genome of *Borrelia recurrentis*, the agent of deadly louse-borne relapsing fever, is a degraded subset of tick-borne *Borrelia duttonii*. *PLoS Genet* **4:** e1000185. doi: 10.1371/journal.pgen.1000185.
Ormsbee RA, Peacock MG, Gerloff R, Tallent G, Wike D. 1978. Limits of rickettsial infectivity. *Infect Immun* **19:** 239–245.
Pan X, Luhrmann A, Satoh A, Laskowski-Arce MA, Roy CR. 2008. Ankyrin repeat proteins comprise a diverse family of bacterial type IV effectors. *Science* **320:** 1651–1654.
Pizarro-Cerda J, Cossart P. 2006. Bacterial adhesion and entry into host cells. *Cell* **124:** 715–727.
Polevoda B, Sherman F. 2007. Methylation of proteins involved in translation. *Mol Microbiol* **65:** 590–606.
Raoult D, Roux V. 1999. The body louse as a vector of reemerging human diseases. *Clin Infect Dis* **29:** 888–911.
Raoult D, Ogata H, Audic S, Robert C, Suhre K, Drancourt M, Claverie JM. 2003. *Tropheryma whipplei* twist: A human pathogenic actinobacteria with a reduced genome. *Genome Res* **13:** 1800–1809.
Raoult D, Dutour O, Houhamdi L, Jankauskas R, Fournier PE, Ardagna Y, Drancourt M, Signoli M, La VD, Macia Y, et al. 2006. Evidence for louse-transmitted diseases in soldiers of Napoleon's Grand Army in Vilnius. *J Infect Dis* **193:** 112–120.
Renesto P, Azza S, Dolla A, Fourquet P, Vestris G, Gorvel JP, Raoult D. 2005. *Rickettsia conorii* and *R. prowazekii* proteome analysis by 2DE-MS: A step toward functional analysis of *Rickettsial* genomes. *Ann N Y Acad Sci* **1063:** 90–93.
Renesto P, Samson L, Ogata H, Azza S, Fourquet P, Gorvel JP, Heinzen RA, Raoult D. 2006. Identification of two putative rickettsial adhesins by proteomic analysis. *Res Microbiol* **157:** 605–612.
Rodionov AV, Eremeeva ME, Balayeva NM. 1991. Isolation and partial characterization of the Mr 100 kD protein from *Rickettsia prowazekii* strains of different virulence. *Acta Virol* **35:** 557–565.
Rosenberg SM. 2001. Evolving responsively: Adaptive mutation. *Nat Rev Genet* **2:** 504–515.
Tamas I, Wernegreen JJ, Nystedt B, Kauppinen SN, Darby AC, Gomez-Valero L, Lundin D, Poole AM, Andersson SG. 2008. Endosymbiont gene functions impaired and rescued by polymerase infidelity at poly(A) tracts. *Proc Natl Acad Sci* **105:** 14934–14939.
Tatusov RL, Koonin EV, Lipman DJ. 1997. A genomic perspective on protein families. *Science* **278:** 631–637.
Turco J, Winkler HH. 1982. Differentiation between virulent and avirulent strains of Rickettsia prowazekii by macrophage-like cell lines. *Infect Immun* **35:** 783–791.
Turco J, Winkler HH. 1994. Cytokine sensitivity and methylation of lysine in *Rickettsia prowazekii* EVir and interferon-resistant *R. prowazekii* strains. *Infect Immun* **62:** 3172–3177.
Vallenet D, Labarre L, Rouy Z, Barbe V, Bocs S, Cruveiller S, Lajus A, Pascal G, Scarpelli C, Medigue C. 2006. MaGe: A microbial genome annotation system supported by synteny results. *Nucleic Acids Res* **34:** 53–65.
Zhang JZ, Hao JE, Walker DH, Yu XJ. 2006. A mutation inactivating the methyltransferase gene in avirulent Madrid E strain of *Rickettsia prowazekii* reverted to wild type in the virulent revertant strain Evir. *Vaccine* **24:** 2317–2323.
Zinsser H. 1935. *Rats, lice, and history*. Broadway House, London, UK.

Received November 26, 2009; accepted in revised form February 11, 2010.

Remerciements

Tout d'abord je tiens à remercier le Pr Didier Raoult de m'avoir accueilli au sein de son laboratoire ainsi que d'avoir accepté de reprendre la direction de ma thèse au bout d'un an et demi. Merci pour votre passion pour la recherche, votre enthousiasme à toute épreuve et énergie débordante qui malgré une pression assez forte, sont restés des moteurs à l'avancement de mes travaux. Ce fut un plaisir de passer ma thèse à vos cotés même si cela n'a pas été facile tous les jours. Mais comme vous le dites si bien : « *Lorsque c'est facile ce n'est pas intéressant*. »

Ensuite je tiens à remercier les membres du jury pour avoir accepté de juger mes travaux de thèse. Merci au Pr Jean-Louis Mège d'avoir accepté de présider le jury mais aussi pour votre disponibilité sans faille envers les étudiants du laboratoire. Merci aux Professeurs Max Maurin et Marie-Laure Joly-Guillou d'avoir accepté de faire parti de mon jury de thèse en qualité de rapporteur ainsi que le Docteur Patricia Renesto en qualité d'examinateur.

Je voudrais remercier aussi toutes les personnes qui ont participé de près ou de loin à cette thèse.

D'abord, je voulais adresser encore un grand merci à Patricia Renesto. Même si nos chemins se sont séparés au bout d'un an et demi je voulais te remercier pour ta gentillesse et ta disponibilité à mon égard ainsi que pour ta rigueur scientifique que j'ai essayé tant bien que mal de recopier. J'espère sincèrement que nos chemins se recroiseront et que nous aurons l'occasion de retravailler ensemble.

Merci à Clarisse et My-Van et donc Patricia (notre chef) qui ont constitué avec moi pendant un bon moment l'équipe Microarray avant qu'elles ne me quittent pour d'autres horizons. J'ai eu énormément de plaisir à travailler à vos cotés même si cela fut beaucoup trop bref à mon goût. Merci à mes collaborateurs de Sophia Antipolis à savoir d'abord M. Barbry et Kevin qui m'ont fourni les microarray, outils indispensables à cette thèse et ensuite le Docteur Richard Thiery, pour m'avoir permis grâce au projet VIACOX d'avoir un financement pour cette thèse ainsi que de m'avoir permis de m'aventurer régulièrement à Sophia Antipolis pour nos réunions semestrielles.

Merci aux chercheurs, médecins, ingénieurs, aux techniciens, aux secrétaires et étudiants que j'ai côtoyés pendant ces 4 ans à la fois aux Rickettsies mais aussi à la Timone.

Un merci avec une mention spéciale à Guy, Jean-Yves et Marie-Laure, M. Hoest et toute l'équipe du P3 qui m'ont initié à la culture des bactéries intracellulaires et au travail en P3. Merci pour toute l'aide que vous m'avez apportée. Je n'aurais jamais pu mener à bien cette thèse sans vous.

Merci à Catherine et toute l'équipe de la Génopôle et ses nombreux stagiaires qui m'ont permis de ne pas rester enclavé dans ma petite salle ARN du 3ème étage et pour toutes les conversations constructives que nous avons échangées.

Un grand merci aussi à tous les étudiants que j'ai pu croiser durant cette thèse et spécialement mes copines de bureau (que j'ai souvent embêtées) : Yassina, Vicky, Claire, Nawal. Merci à Nicolas, Fabrice, Sophie, Pepette, Nada, Meriem, Diane et Manolis. Merci aux autres étudiants, la liste est beaucoup trop longue pour tous vous citer donc entre autres, Kalioppi, Saru, Angélique, Isabelle, Mano, Sudhir, Amélie, Malgo…. (Désolé pour ceux qui ne sont pas cités vous êtes beaucoup trop).

Merci à mes deux camarades de déjeuner Thi-Phong et Thi-Tien. Grâce à vous j'ai eu le plaisir de manger pratiquement tous les jours avec les deux plus belles filles de la fac. J'espère que nos chemins ne se sépareront pas. Vive la Viet-connexion.

Je tiens aussi à faire un petit flash-back nostalgie revival en remerciant mes profs de Sciences Naturelles (M. Sahors, Mme Auvray, Mme Mathé) et aussi mes profs à l'université (Bénédicte, Francis et tant d'autres). Ce sont ces gens, passionnés par leurs métiers et à l'écoute de leurs étudiants, qui m'ont donné ce goût pour la biologie. J'ai eu de la chance et l'honneur de les croiser pendant mon parcours. Je leur dédie cette thèse.

Une merci un peu décalé à l'inventeur des sites internet VDM et TDM qui m'a permis à moi comme à beaucoup d'autres thésards de relativiser les problèmes que l'on peu rencontrer lorsque l'on est doctorant.

Merci à ma binôme préférée (et sa petite famille) qui a réussi à me faire venir à Marseille, même si je lui avais toujours juré que je n'y mettrai jamais un pied. Merci à mes amis et aux potes de l'USRO. Ils m'ont permis à chaque fois que je suis retourné à Paris de décompresser et de me ressourcer. J'espère revenir très bientôt parmi vous.

Merci à mes parents et mon frère pour le support, l'amour et l'attention qu'ils me portent. Cette thèse est aussi la votre. Je vous aime.

Enfin un merci spécial à une personne qui se reconnaitra en lisant ces remerciements. Merci pour tout ;)

Voilà déjà 4 ans que cette thèse m'a éloigné de chez moi et de mes racines. Malgré une adaptation difficile et une envie continuelle de retourner à Paris, j'ai eu le plaisir de côtoyer des gens venus d'horizons tellement différents. Cette aventure qu'a constitué ma thèse ne m'a pas seulement apporté un diplôme mais m'a également fait grandir à vos cotés. Certains d'entre vous ont une place particulière dans mon cœur.

Je voulais finir mes remerciements par une citation mais n'ayant pu en trouver une qui pouvait résumer ces 4 années j'en ai donc choisi plusieurs d'hommes plus ou moins célèbre qui, je l'espère, vous ferons un peu découvrir ce que l'on peut vivre au cours d'une thèse.

" Un peu de foi éloigne de Dieu, beaucoup de science y ramène. "

Francis Bacon

" Le commencement de toutes les sciences c'est l'étonnement de ce que les choses sont ce qu'elles sont "

Aristote

" En Science, la phrase la plus excitante que l'on peut entendre, celle qui annonce de nouvelles découvertes, ce n'est pas Eurêka mais c'est drôle. "

Isaac Asimo

" Je n'ai jamais rencontré d'homme si ignorant qu'il n'eut rien à m'apprendre. "

Galilée

" Il y a plus de courage que de talent dans la plupart des réussites. "

Félix Leclerc

" N'acceptez jamais la défaite, vous êtes peut être à un pas de la réussite. "

Jack E Addington

" Chaque publication scientifique ne sert qu'à poser 10, 20 questions. Chaque découverte scientifique est passionnante parce qu'elle ouvre un univers de questions. Si les questions vous angoissent, ne soyez pas scientifique. "

Boris Cyrulnik

" Vers l'infini et au delà. "

Buzz l'Eclair